DRAFT VERSION APRIL 26, 2013
Preprint typeset using LaTeX style emulateapj v. 08/22/09

THE COLUMN DENSITY DISTRIBUTION AND CONTINUUM OPACITY OF THE INTERGALACTIC AND CIRCUMGALACTIC MEDIUM AT REDSHIFT $\langle Z\rangle = 2.4$

GWEN C. RUDIE,[2] CHARLES C. STEIDEL,[2] ALICE E. SHAPLEY,[3] & MAX PETTINI[4]

Draft version April 26, 2013

ABSTRACT

We present new high-precision measurements of the opacity of the intergalactic and circumgalactic medium (IGM, CGM) at $\langle z\rangle = 2.4$. Using Voigt profile fits to the full Lyman α and Lyman β forests in 15 high-resolution high-S/N spectra of hyperluminous QSOs, we make the first statistically robust measurement of the frequency of absorbers with H I column densities $14 \lesssim \log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 17.2$. We also present the first measurements of the frequency distribution of H I absorbers in the volume surrounding high-z galaxies (the CGM, 300 pkpc), finding that the incidence of absorbers in the CGM is much higher than in the IGM. In agreement with Rudie et al. (2012), we find that there are fractionally more high-$N_{\rm HI}$ absorbers than low-$N_{\rm HI}$ absorbers in the CGM compared to the IGM, leading to a shallower power law fit to the CGM frequency distribution. We use these new measurements to calculate the total opacity of the IGM and CGM to hydrogen-ionizing photons, finding significantly higher opacity than most previous studies, especially from absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$. Reproducing the opacity measured in our data as well as the incidence of absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$ requires a broken power law parameterization of the frequency distribution with a break near $N_{\rm HI} \approx 10^{15}$ cm^{-2}. We compute new estimates of the mean free path ($\lambda_{\rm mfp}$) to hydrogen-ionizing photons at $z_{\rm em} = 2.4$, finding $\lambda_{\rm mfp} = 147 \pm 15$ Mpc when considering only IGM opacity. If instead, we consider photons emanating from a high-z star-forming galaxy and account for the local excess opacity due to the surrounding CGM of the galaxy itself, the mean free path is reduced to $\lambda_{\rm mfp} = 121 \pm 15$ Mpc. These $\lambda_{\rm mfp}$ measurements are smaller than recent estimates and should inform future studies of the metagalactic UV background and of ionizing sources at $z \approx 2-3$.

Subject headings: intergalactic medium — quasars: absorption lines

1. INTRODUCTION

Observations of the Lyman α forest imprinted upon the spectra of background QSOs provide a rare window into the nature and evolution of baryonic structures in the universe. With the advent of the theory of cold dark matter (CDM), it was recognized that self-gravitating halos could be the sites of uncollapsed Lyα clouds (Umemura & Ikeuchi 1985; Rees 1986). In the modern version of this model, the Lyα forest simply represents the array of density fluctuations present in the early universe and evolved under gravity to the present day (Tytler 1987; Cen et al. 1994; Hernquist et al. 1996), modulo the effect of the formation of stars, black holes, and galaxies which appear to have polluted a large fraction of the cosmic hydrogen with metals, even at early times (Songaila & Cowie 1996; Davé et al. 1998; Schaye et al. 2003; Simcoe et al. 2004; Ryan-Weber et al. 2009; Becker et al. 2009; Simcoe 2011; Simcoe et al. 2011; Becker et al. 2011, 2012) and led to the ionization of the vast majority of the gas (Gunn & Peterson 1965).

There is an extensive literature concerned with the measurement of cosmological parameters using observations of the Lyα forest [for a review, see Rauch (1998) and Meiksin (2009)]. Considerable progress has been made in the last several decades in measuring the detailed properties of the IGM and their implications for cosmology, structure formation, reionization, and the metagalactic background.

One of the most fundamental of these measurements is the number of absorbers as a function of column density ($N_{\rm HI}$) and redshift that can be seen along a given path through the Universe. This measurement has required significant observational effort, in a large part due to the very large dynamic range of the measurements ($10^{12} \lesssim N_{\rm HI} \lesssim 10^{22}$ cm^{-2}) which have necessitated surveys optimized for particular ranges in $N_{\rm HI}$. This requirement is easily understood by considering the curve of growth of the Lyα line of neutral hydrogen.

The most common absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 14$ are optically thin in the Lyman α transition and are thus on the linear part of the curve of growth. As such, their properties can be measured with a few high-resolution ($\sim$10 km s^{-1}) high-signal to noise ratio (S/N $\sim$ 50) spectra (Kim et al. 2002). At the highest column densities, $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 20.3$, the Lorentzian wings of the absorption line in Lyα become obvious against the continuum of the background source. Such rare pockets of predominantly neutral hydrogen, commonly referred to as Damped Lyman α absorbers (DLAs), can be identified and measured to high statistical precision with a large number of low to moderate resolution spectra of background QSOs as has been done using (e.g.) the QSO sample within the Sloan Digital Sky Survey (Noterdaeme et al. 2009).

Absorbers with intermediate $N_{\rm HI}$, $14 \lesssim \log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 20$, lie on the flat part of the curve of growth thus requiring different techniques. Systems with $N_{\rm HI} > 10^{17.2}$ cm^{-2} have

Electronic address: gwen@astro.caltech.edu
[1] Based on data obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and the National Aeronautics and Space Administration, and was made possible by the generous financial support of the W.M. Keck Foundation.
[2] Cahill Center for Astronomy and Astrophysics, California Institute of Technology, MS 249-17, Pasadena, CA 91125, USA
[3] Department of Astronomy, University of California, Los Angeles, 430 Portola Plaza, Los Angeles, CA 90024, USA
[4] Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

an optical depth to hydrogen-ionizing photons $\tau \geq 1$, and so may be recognized by the strong breaks they produce in background QSO spectra at the Lyman limit (912Å) in their rest frame. These so-called Lyman Limit Systems (LLS) can thus be easily discovered using low resolution spectra; however, a large sample is needed because they are comparatively rare. For studies at $z \lesssim 2.6$, surveys of LLSs require space-based observations as the Lyman limit shifts below the atmospheric cutoff at UV wavelengths. However, over the last 30 years with a combination of ground and space based observations, the distribution of LLSs at $2 < z < 3$ has been well characterized (Tytler 1982; Sargent et al. 1989; Stengler-Larrea et al. 1995; Ribaudo et al. 2011; O'Meara et al. 2012).

Finally there is the historically most-problematic range, $14 \lesssim \log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 17.2$. These intermediate-$N_{\rm HI}$ absorbers are saturated in Lyα but have relatively low opacity to hydrogen-ionizing photons and therefore weak "Lyman breaks." For this reason, observational constraints on their statistical incidence have been approximate at best.

In this paper, we address these intermediate-$N_{\rm HI}$ absorbers using Voigt profile fits to all-available higher-order Lyman series transitions. We take advantage of the increased dynamic range of the measurements afforded by the decreased oscillator strength of the Lyβ,γ, etc. transitions to accurately measure the frequency distribution of these absorbers, $f(N,X)$, for the first time in a statistically robust sample. In our derivation of an analytic representation of $f(N,X)$ we place emphasis on finding a method which best reproduces the measured opacity of absorbers to hydrogen ionizing sources.

These new measurements are crucial to our understanding of the opacity of the universe to hydrogen-ionizing photons. While individually absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ have relatively low optical depths at their Lyman limit, they are much more common than LLSs, and therefore their ensemble contributes nearly half of the IGM's opacity. Here, we present the most precise measurements of the continuum opacity of absorbers that are optically thin at the Lyman limit.

In this paper, we also attempt to account for the fact that galaxies are located in regions of relatively high gas density. Rudie et al. (2012) clearly demonstrated that typical galaxies at $z \approx 2.3$ have large quantities of H I surrounding them to 300 physical kpc (pkpc), and that there is excess H I compared to random places in the IGM to $\sim$2 physical Mpc (pMpc). In the case that galaxies such as those studied in Rudie et al. (2012) are significant contributors to the ionizing background, the effect of their surrounding circumgalactic gas distribution is an important consideration. In this work, we attempt to separate the probability that an ionizing photon will escape the ISM of its galaxy from the probability that such a photon, after escaping to 50 pkpc, will also escape the enhanced opacity found in its circumgalactic medium (CGM).

We begin in Section 2 with a description of the unique data that enable these very precise $N_{\rm HI}$ measurements. In Section 3 we present the new $N_{\rm HI}$ frequency distribution measurements. Section 4 considers the frequency distribution within the CGM of galaxies and Section 5 quantifies the total opacity of the IGM and the environment of the source galaxy. Updated measurements of the mean free path ($\lambda_{\rm mfp}$) to hydrogen-ionizing photons are presented in Section 6 along with their implications. The paper is summarized in Section 7.

Throughout this paper we assume a Λ-CDM cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_{\rm m} = 0.3$, and $\Omega_\Lambda = 0.7$. All distances are expressed in physical (proper) units unless stated otherwise. We use the abbreviations pkpc and pMpc to indicate physical units. At the path-weighted mean redshift of the absorber sample ($\langle z \rangle = 2.37$), the age of the universe is 2.7 Gyr, the look-back time is 10.7 Gyr, and 8.2 pkpc subtends one arcsecond on the sky.

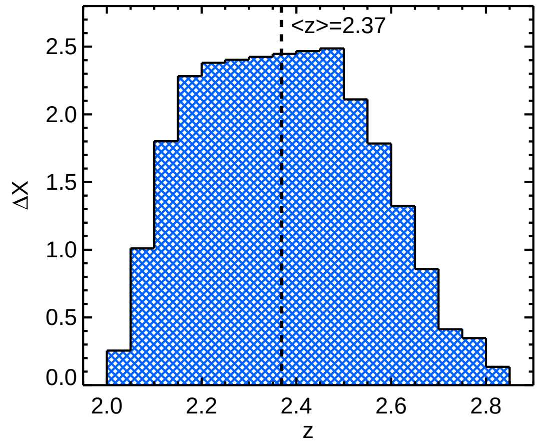

Figure 1. Histogram of the pathlength sampled as a function of redshift. The path-length-weighted average redshift of the sample is $z = 2.37$.

2. DATA AND ANALYSIS

The data presented in this paper constitute a subset of the Keck Baryonic Structure Survey (KBSS). The KBSS was designed to detect and characterize the gaseous distribution surrounding star-forming galaxies at $2 < z < 3$ during the peak of cosmic star formation and black hole growth. The data include rest-frame UV spectra of 2188 star-forming galaxies located in 15 fields surrounding the lines of sight to hyperluminous ($m_V \simeq 15.5-17$) QSOs. The present work focuses primarily on the H I absorption line statistics resulting from analysis of high-resolution, high signal-to-noise echelle spectra of the 15 KBSS QSOs. A full description of the KBSS will be presented by Steidel et al. (in prep), but the portions of the data most relevant to this work (as well as a detailed description of the analysis of the QSO data) can be found in Rudie et al. (2012). Here, we provide a brief description.

The 15 KBSS QSOs were observed primarily with the High Resolution Echelle Spectrometer (HIRES; Vogt et al. 1994) on the Keck I telescope. The HIRES spectra have $R \simeq 45,000$ (FWHM$\simeq$ 7 km s^{-1}), S/N $\sim 50-200$ per pixel, and cover at least the wavelength range 3100 – 6000 Å with no spectral gaps. The UV/blue wavelength coverage provides a significant advantage over other data sets as it enables the observation of Lyβ λ1025.7 down to at least $z = 2.2$ in all 15 of our sightlines, and to significantly lower redshift in many.[5] The additional constraints provided by Lyβ (and in many cases, additional Lyman series transitions) allow for much more accurate measurements of H I for $N_{\rm HI} = 10^{14}-10^{17}$ cm^{-2} than can be made with spectra including only the Lyα transition. A full discussion of the redshift sensitivity of our fitting procedure is given in Appendix A.

The reduction of the HIRES spectra was performed using T. Barlow's MAKEE package as described in detail by

[5] The minimum wavelength that can be used for line fitting is determined by both the wavelength coverage of the spectrum (set by the spectrograph and the atmosphere) and by the location of the highest-redshift Lyman limit system along each line of sight. The minimum redshift for which we perform line fitting is shown in Table 1.

Table 1
KBSS Absorption Line Sample

Name	RA	Dec	z_{QSO}[a]	λ_{min}[b] [Å]	z range	Δz	ΔX[c]	N_{gal}[d]	S/N Lyα[e]	S/N Lyβ[e]
Q0100+130 (PHL957)	01:03:11.3	+13:16:18.2	2.721	3133	2.0617– 2.6838	0.62	2.02	5	77	50
HS0105+1619	01:08:06.4	+16:35:50.0	2.652	3230	2.1561– 2.6153	0.46	1.50	0	127	89
Q0142–09 (UM673a)	01:45:16.6	–09:45:17.0	2.743	3097	2.0260– 2.7060	0.68	2.21	3	71	45
Q0207–003 (UM402)	02:09:50.7	–00:05:06.5	2.872	3227	2.1532– 2.8339	0.68	2.26	1	82	55
Q0449–1645	04:52:14.3	–16:40:16.2	2.684	3151	2.0792– 2.6470	0.57	1.84	2	73	41
Q0821+3107	08:21:07.6	+31:07:51.2	2.616	3239	2.1650– 2.5794	0.41	1.35	2	50	33
Q1009+29 (CSO 38)	10:11:55.6	+29:41:41.7	2.652[f]	3186	2.1132– 2.6031	0.49	1.59	0	99	58
SBS1217+499	12:19:30.9	+49:40:51.2	2.704	3098	2.0273– 2.6669	0.64	2.07	5	68	38
HS1442+2931	14:44:53.7	+29:19:05.6	2.660	3152	2.0798– 2.6237	0.54	1.76	6	99	47
HS1549+1919	15:51:52.5	+19:11:04.3	2.843	3165	2.0926– 2.8048	0.71	2.35	4	173	74
HS1603+3820	16:04:55.4	+38:12:01.8	2.551[g]	3181	2.1087– 2.5066	0.40	1.28	2	108	58
Q1623+268 (KP77)	16:25:48.8	+26:46:58.8	2.535	3126	2.0544– 2.4999	0.45	174	3	48	28
HS1700+64	17:01:00.6	+64:12:09.4	2.751	3138	2.0668– 2.7138	0.65	2.11	4	98	42
Q2206–199	22:08:52.1	–19:43:59.7	2.573	3084	2.0133– 2.5373	0.52	1.68	3	88	46
Q2343+125	23:46:28.3	+12:48:57.8	2.573	3160	2.0884– 2.5373	0.45	1.45	3	71	45

[a] The redshift of the QSO (Trainor & Steidel 2012)
[b] The minimum wavelength covered in the HIRES QSO spectrum
[c] The comoving pathlength in this line of sight (equation 3).
[d] The number of galaxies in our LRIS survey with impact parameters < 300 pkpc.
[e] The average signal to noise ratio per pixel of the QSO spectrum in the wavelength range pertaining to IGM Lyα and Lyβ absorption.
[f] The redshift of this QSO was revised after the fitting of the HIRES spectrum was completed. The maximum redshift assumed $z_{QSO} = 2.640$
[g] The redshift of this QSO was revised after the fitting of the HIRES spectrum was completed. The maximum redshift assumed $z_{QSO} = 2.542$

Rudie et al. (2012). The analysis presented here includes a Voigt profile fit to the full Lyα and Lyβ forests in these spectra. The redshift range included in the fit is bounded at high redshift by the proximity zone of the QSO (taken to be 3000 km s^{-1}) in which absorbers may be ejected from or ionized by the QSO itself. The low-redshift cut was chosen to ensure that all H I absorbers included in the sample had been observed in both the Lyα and Lyβ transition. The final redshift windows used for each spectrum are detailed in Table 1.

The line fitting was facilitated by a semi-automatic code that estimates input parameters (redshift z, H I column density $N_{\rm HI}$, and Doppler parameter $b_{\rm d}$ of each absorber) using a cross-correlation technique as described in Rudie et al. (2012). These parameters were then input into the χ^2 minimization code VPFIT[6] written by R.F. Carswell and J.K. Webb. The results were iteratively checked, altered, and rerun until a good fit was achieved. Median uncertainties in $\log(N_{\rm HI})$ for these absorber measurements are listed in Table 2. Examples of fits to absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 15.5$ are given in Appendix A along with an assessment of the completeness of the catalog as a function of redshift.

The result of this process is an H I absorber catalog including 5758 absorbers with $12.0 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ and $2.02 < z < 2.84$ over a total redshift pathlength of $\Delta z = 8.27$. This represents the largest Lyα forest catalog to date at these redshifts and increases by an order of magnitude the number of absorbers measured using the additional constraints provided by higher-order Lyman series transitions.

3. $N_{\rm HI}$ FREQUENCY DISTRIBUTION

The differential column density distribution, first defined by Carswell et al. (1984), is a useful way of parameterizing the frequency of absorbers as a function of $N_{\rm HI}$. We consider the differential column density distribution per unit pathlength, $f(N_{\rm HI}, X)$, defined as

$$f(N_{\rm HI},X)dN_{\rm HI}dX = \frac{m}{\Delta N_{\rm HI}\Delta X}dN_{\rm HI}dX \qquad (1)$$

[6] http://www.ast.cam.ac.uk/∼rfc/vpfit.html; © 2007 R.F. Carswell, J.K. Webb

Table 2
Median $\log(N_{\rm HI})$ uncertainty for absorber measurements

$\log(N_{\rm HI}/{\rm cm}^{-2})$	$\sigma_{\log(N_{\rm HI})}$
12.0 – 12.5	0.10
12.5 – 13.0	0.07
13.0 – 13.5	0.07
13.5 – 14.0	0.06
14.0 – 14.5	0.04
14.5 – 15.0	0.03
15.0 – 15.5	0.03
15.5 – 16.0	0.04
16.0 – 16.5	0.07
16.5 – 17.0	0.06
17.0 – 17.5	0.09

where m is the observed number of absorbers with column densities in the range $\Delta N_{\rm HI}$. ΔX is the comoving pathlength of the survey summed over the sightlines. This comoving pathlength was introduced by Bahcall & Peebles (1969) in part to understand if the statistics of absorbers were consistent with having been produced by galaxies. With this formalism, absorbers with a constant physical size and comoving number density will have constant $f(N,X)$. The comoving pathlength of a single sightline, ΔX_i, is defined as:

$$dX = \frac{H_0}{H(z)}(1+z)^2 dz \qquad (2)$$

$$\Delta X_i = \int_{z_{min}}^{z_{max}} dX = \int_{z_{min}}^{z_{max}} \frac{(1+z)^2}{\sqrt{\Omega_\Lambda + \Omega_{\rm m}(1+z)^3}}dz \qquad (3)$$

The total pathlength covered by this survey is $\Delta X = 26.9$ and the ΔX covered as a function of redshift summed over the 15 lines of sight is shown in Figure 1. The pathlength-weighted mean redshift of the sample is $\langle z \rangle = 2.37$. Measurements of $f(N,X)$ derived from the KBSS sample of absorbers with $12.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ are presented in Figure 2. The data points are located at the mean $\log(N_{\rm HI})$ of the bin edges.

3.1. *Parameterizing the frequency distribution*

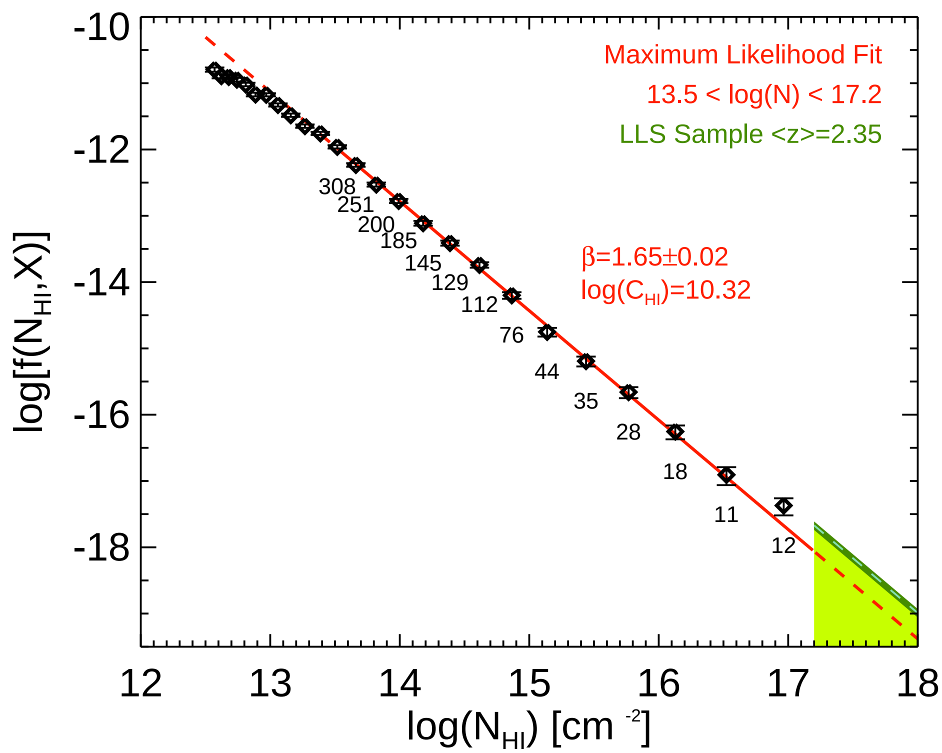

Figure 2. The frequency distribution of absorbers from the KBSS H I sample (black data points). For bins with $N_{\rm HI} > 13.5$ cm^{-2}, the number of absorbers per bin is displayed. The (red) line is a maximum likelihood power-law fit to the distribution of absorbers with $13.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$. The (green) triangle represents the frequency of LLS calculated from other studies as discussed in §3.3. The dark green band at the top of the triangle represents the error in this estimation. Note this is a single power law fit to the data. Section 5.1 discusses broken-power law fits which better reproduce some aspects of the data, especially the incidence of LLSs.

It has become standard to represent $f(N_{\rm HI},X)$ by a power law, or a series of broken power laws of the form:

$$f(N_{\rm HI},X)dN_{\rm HI}dX = C_{\rm HI}N_{\rm HI}^{-\beta}dN_{\rm HI}dX. \quad (4)$$

Tytler (1987) first noted that the full column density distribution from $12 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 21$ was reasonably well approximated by a single power law with $\beta = 1.5$. The detailed deviations from a single power law at various $N_{\rm HI}$ and the change in normalization and slope of $f(N_{\rm HI},X)$ as a function of redshift have been considered by numerous authors (e.g., Sargent et al. 1989; Petitjean et al. 1993; Davé & Tripp 2001; Kim et al. 2002; Penton et al. 2004; Janknecht et al. 2006; Lehner et al. 2007; Prochaska et al. 2010; Ribaudo et al. 2011; O'Meara et al. 2012).

The most notable deviations from a single power law appear at the ends of the distribution. The low-$N_{\rm HI}$ end appears to flatten, an effect generally attributed to a combination of incompleteness due to line blending and blanketing and a true physical turnover (Hu et al. 1995; Ellison et al. 1999; Schaye 2001; Kim et al. 2002). The deviation from a single power-law for $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 19$ is believed to be due to the transition of gas from optically thin to self-shielding regimes (Murakami & Ikeuchi 1990; Petitjean et al. 1992; Katz et al. 1996; Corbelli et al. 2001; Zheng & Miralda-Escudé 2002; Altay et al. 2011; McQuinn et al. 2011). The most recent estimates of the frequency distribution of absorbers with $19 < \log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 20.3$ at $z \approx 2$ are by O'Meara et al. (2007); the frequency distribution of damped Lyman α systems with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 20.3$ is measured to relatively high precision using the Sloan Digital Sky Survey (SDSS) QSOs by Noterdaeme et al. (2009). In this paper, we leverage the KBSS QSO absorption line data set to measure $f(N,X)$ for absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$.

3.2. *Maximum Likelihood Method*

A statistically rigorous estimate of the power law exponent and normalization can be found via the Maximum Likelihood Method. The maximum likelihood estimator (MLE) for the power law exponent is:

$$\hat{\beta} = 1 + n\left[\sum_{i=1}^{n} \ln\frac{N_i}{N_{\rm min}}\right]^{-1} \quad (5)$$

where n is the total number of absorbers in the sample and $N_{\rm min}$ in the column density of the lowest-column absorber (see e.g., Clauset et al. 2007). The standard error on $\hat{\beta}$, derived from the width of the likelihood function, is:

$$\sigma_{\hat{\beta}} = \frac{\hat{\beta}-1}{\sqrt{n}} \quad (6)$$

Similarly, the estimator of the normalization, $C_{\rm HI}$, is

$$\hat{C}_{\rm HI} = \frac{n(1-\hat{\beta})}{\left(N_{\rm max}^{(1-\hat{\beta})} - N_{\rm min}^{(1-\hat{\beta})}\right)\Sigma\Delta X} \quad (7)$$

(Tytler 1987) where $\Sigma\Delta X$ is the sum of the pathlength over

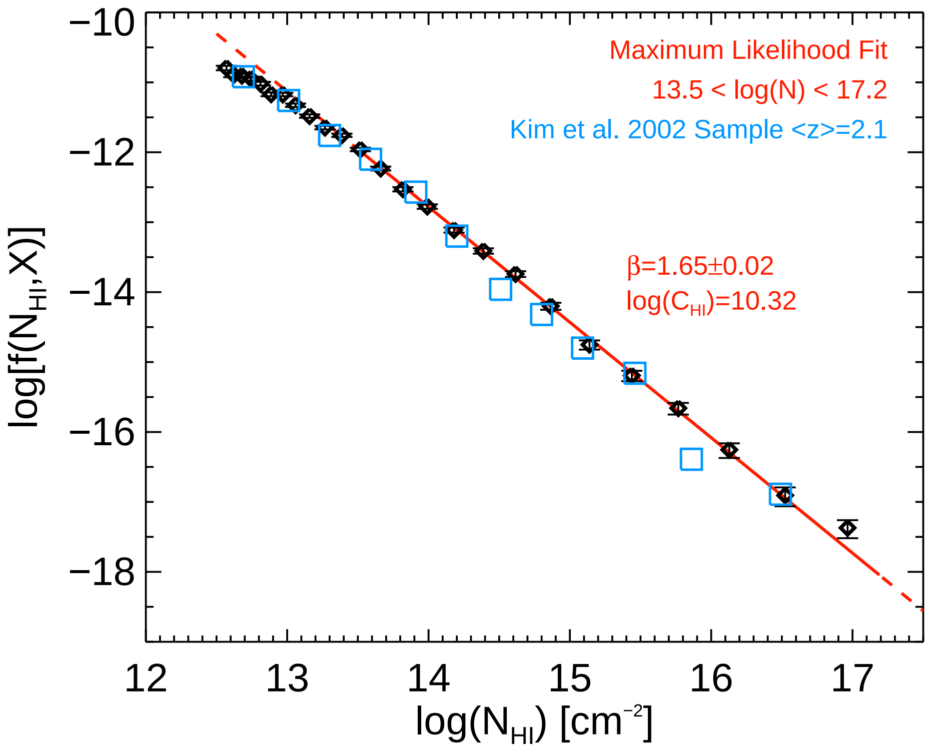

Figure 3. The frequency distribution of absorbers from the KBSS H I sample (black points) and the maximum likelihood fit to the data (red curve). For comparison, over plotted in (blue) squares are the data points from the $\langle z \rangle = 2.1$ sample from Kim et al. (2002). Because the analysis of the KBSS sample includes higher-order H I transitions, we can measure the frequency of absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 14$ with much higher fidelity than has been previously possible. Notably, there is no evidence for a strong break in the power law distribution near $\log(N_{\rm HI}/{\rm cm}^{-2}) \approx 14$ as has been previously suggested.

Table 3
Maximum Likelihood Fits[a]

$\log(N_{\rm HI,min})$	$\hat{\beta}$	$\sigma_{\hat{\beta}}$	$\log(\hat{C}_{\rm HI})$
12.5	1.518	0.008	8.43
12.6	1.544	0.008	8.80
12.7	1.567	0.009	9.13
12.8	1.585	0.010	9.38
12.9	1.605	0.011	9.67
13.0	1.616	0.012	9.82
13.1	1.627	0.013	9.99
13.2	1.636	0.014	10.11
13.3	1.651	0.015	10.33
13.4	1.658	0.016	10.44
13.5	1.650	0.017	10.32
13.6	1.636	0.018	10.11
13.7	1.639	0.020	10.15
13.8	1.645	0.021	10.25
13.9	1.649	0.023	10.31
14.0	1.654	0.025	10.39

[a] Maximum likelihood fits to the data with $\log(N_{\rm HI,min}) \leq \log(N_{\rm HI}/{\rm cm}^{-2}) \leq 17.2$.

all 15 sightlines .

In Figure 2 the MLE fit to the data is shown for absorbers with $13.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$. Note that a single power law reproduces the distribution with reasonable fidelity. No strong breaks in the distribution are evident.

The choice of a minimum $N_{\rm HI}$= 13.5 used in the fit is motivated as follows. As shown in Figure 2, $f(N_{\rm HI},X)$ for $\log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 13.5$ appears to flatten. As mentioned above, this flattening has previously been observed (Hu et al. 1995; Ellison et al. 1999; Kim et al. 2002) and is generally attributed to a combination of incompleteness in the sample caused by line blending and a true physical turnover in $f(N_{\rm HI},X)$. Here we do not attempt to diagnose the cause of the low-$N_{\rm HI}$ turnover, but simply note that above $\log(N_{\rm HI}/{\rm cm}^{-2}) \geq 13.5$, $f(N,X)$ appears to be well fit by a single power law. More quantitatively, one can measure the change in power law slope introduced by varying the minimum value of $N_{\rm HI}$ included in the fit as shown in Table 4. Note that a choice of $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 13.3$ produces insignificant change in the maximum likelihood estimate of the slope, $\hat{\beta}$.

Prior to the present work, the largest existing sample of Lyα forest absorbers at these redshifts was described by Kim et al. (2002). Their sample included 2130 H I absorbers with $12.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17$ observed along 9 lines of sight. The most marked improvement over the previous measurements provided by our study is the use of higher-order transitions of H I which were generally not included in previous work. These allow for more precise measurements of $N_{\rm HI}$ above the saturation point of Lyα, $N_{\rm HI} \approx 10^{14-14.5}$ cm^{-2}. Above $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 15.5$ where the Lyβ transition saturates, the sample presented herein may suffer a small degree of incompleteness at $z \lesssim 2.4$ as discussed in Appendix A.

In Figure 3, the KBSS data (black diamonds) are compared with the measurements of Kim et al. (2002) at $\langle z \rangle = 2.1$ (blue squares). Notably, at $\log(N_{\rm HI}/{\rm cm}^{-2}) < 14$ there is good agreement between the two data sets, but for $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 14$, the Kim et al. (2002) sample exhibits significant scatter not present in the KBSS measurements, with a systematic

tendency to under-estimate the number of intermediate-$N_{\rm HI}$ ($14 \lesssim \log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 17$) absorbers compared to our measurements. Potential incompleteness found in the lower-z portion of the catalog presented herein (see Appendix A) may also affect the Kim et al. (2002) catalog for absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 14$ and may explain why their measurements are systematically lower than those presented here.

3.3. *Lyman Limit Systems*

In this section, we quantify the incidence of absorbers with $N_{\rm HI} > 10^{17.2}$ cm^{-2}, Lyman Limit Systems (LLSs) which are optically thick to hydrogen-ionizing photons. Because LLSs are relatively rare, and because their $N_{\rm HI}$ are challenging to determine accurately via line fitting, we use samples reported in the literature to measure the frequency of optically thick absorbers at the same redshifts as the KBSS sample.

We estimated the incidence of LLSs with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$ at $\langle z \rangle \sim 2.4$ using data from two surveys that nicely bracket the relevant range of redshifts for our IGM sample. We evaluated the LLSs over a somewhat wider redshift range than $2 \lesssim z \lesssim 2.8$ (but having the same mean redshift) in order to mitigate the statistics of small numbers. Over the range $1.65 < z < 2.55$, we used data from Ribaudo et al. (2011) from their survey compilation of archival HST spectroscopy; results for $2.70 < z < 2.95$ are based on a subset of the ground-based surveys of Sargent et al. (1989) and Stengler-Larrea et al. (1995). The sub-samples are roughly equivalent in their statistical significance (comparable total pathlength and number of LLSs), with path-weighted mean redshifts of $\langle z \rangle = 2.08$ and $\langle z \rangle = 2.84$, respectively. The path-weighted mean redshift for the full sample is $\langle z_{\rm LLS} \rangle = 2.41$ based on 46 LLSs with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$ in a total redshift path of $\Delta z = 25.95$ ($\Delta X = 88.31$). These LLS samples provide a maximum likelihood estimate of the number of LLS per unit pathlength, $dn_{\rm LLS}/dX = 0.52 \pm 0.08$.

A recent study of LLS near the redshift range of interest ($2.0 < z < 2.6$) using uniform HST data is described in O'Meara et al. (2012). They measure an integral constraint on the number of LLS absorbers with $\tau_{912} > 1, \log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$ and $\tau_{912} > 2, \log(N_{\rm HI}/{\rm cm}^{-2}) > 17.5$, where τ_{912} is the optical depth at the Lyman Limit, 912Å, in the absorber's rest frame. These authors found $dn_{\tau>1}/dX = 0.48^{+0.09}_{-0.12}$, in excellent agreement with our compilation described above ($dn_{\rm LLS}/dX = 0.52 \pm 0.08$). For more optically thick systems, O'Meara et al. (2012) found $dn_{\tau>2}/dX = 0.28^{+0.06}_{-0.06}$. The consistency of these measurements with the lower-$N_{\rm HI}$ frequency distribution will be discussed further in §5.

Unfortunately, surveys that rely on detecting breaks in the spectrum of QSOs at the redshift of the LLS can rarely measure $N_{\rm HI}$ accurately, especially for absorbers with $\tau_{912} > 2$. As such, a measurement of the power law index of $f(N,X)$ in the LLS regime typically relies on the combination of integral constraints (dn/dX) for the LLS, subDLA, and DLA regimes, as well as the measured values of $f(N,X)$ for Lyα forest absorbers. As we now have much more statistically robust measurements of β for Lyα forest absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 17.2$, we re-examine the power law fit to $f(N,X)$ in the LLS regime.

Here, we test the consistency of the measured incidence of LLSs with the extrapolated slope inferred from the MLE method for absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$. First, we note that LLS surveys are sensitive to all absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$. Observations of DLAs (e.g., Noterdaeme et al. 2009) show that there is a strong turnover in the frequency of DLAs at $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 21$, suggesting there are very few absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 22$. Here, we assume that $dn_{\rm LLS}/dX$ is sensitive to absorbers with $17.2 < \log(N_{\rm HI}/{\rm cm}^{-2}) > 22.0$.

Next, we consider the incidence of LLSs that would be found if the MLE power law fit constrained by our data at $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ held for the entire LLS regime. We calculate the incidence of LLSs given the fit parameters as follows:

$$dn_{\rm LLS}/dX = \int_{N_{\rm min}}^{N_{\rm max}} f(N_{\rm HI}, X) dN_{\rm HI} \quad (8)$$

$$dn_{\rm LLS,fit}/dX = \int_{N_{\rm min}}^{N_{\rm max}} C_{\rm HI} N_{\rm HI}^{-\beta} dN_{\rm HI}. \quad (9)$$

where $\log(N_{\rm min}/{\rm cm}^{-2}) = 17.2$ and $\log(N_{\rm max}/{\rm cm}^{-2}) = 22.0$. We find that this extrapolation to $\log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$ underestimates the number of LLSs significantly giving $dn_{\rm LLS,fit}/dX = 0.21$ (compared to $dn_{\rm LLS,data}/dX = 0.52 \pm 0.08$).

The green triangle in Figure 2 represents a similar exercise graphically on the $f(N,X)$ plot. If we assume that the MLE determination of β for the forest holds in the LLS regime, we can calculate the value of $C_{\rm HI,LLS}$ implied by the measured $dn_{\rm LLS}/dX$ by inverting equation 9. Then the values of $f(N,X)$ suggested by the data would be given by:

$$f(N_{\rm p}, X) = C_{\rm HI,LLS} N_{\rm p}^{-\beta}. \quad (10)$$

for all values of $N_{\rm p}$. The height of the green triangle thus represents the values of $f(N,X)$ if $\beta_{\rm MLE}$ holds of the LLS regime. The dark green band at the top of the triangle is an estimate in the uncertainty of this normalization due to the uncertainty in $dn_{\rm LLS}/dX$. Again, we can see clearly that the MLE fit to the frequency distribution (red dotted line) significantly underpredicts the measured LLS incidence.

We have seen that a single power law well approximates the intermediate-$N_{\rm HI}$ absorbers; however, its extrapolation to $\log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$ underestimates the number of LLSs significantly giving $dn_{\rm LLS,fit}/dX = 0.21$. In §5, we discuss other methods for parameterizing the distribution which better reproduce the high-$N_{\rm HI}$ end of the frequency distribution.

4. FREQUENCY DISTRIBUTIONS WITH RESPECT TO THE POSITIONS OF GALAXIES

Rudie et al. (2012) considered the distribution of H I with respect to the positions of star-forming galaxies with $2.0 < z < 2.8$ using the absorption line data presented here. The study explored the kinematic and geometric structure of the CGM, concluding that the majority of the strong excess absorption is found within 300 physical kpc (pkpc) of galaxies in the transverse direction and within 300 km s^{-1} of the systemic redshift of galaxies; however, a tail of excess H I absorbers continues to 2 pMpc and 700 km s^{-1}.

The KBSS spectroscopic galaxy sample consist of UV-color selected galaxies with UV luminosities $0.25 < L/L^* < 3.0$ (Reddy & Steidel 2009) and a typical bolometric luminosity of $\sim 2.5 \times 10^{11} L_\odot$ (Reddy et al. 2008, 2012). They have star formation rates of $\sim 30\ M_\odot$ yr^{-1} (Erb et al. 2006) and halo masses of $\sim 10^{12}\ M_\odot$ (Adelberger et al. 2005; Conroy et al. 2008; Trainor & Steidel 2012, O. Rakic et al. 2012 in preparation).

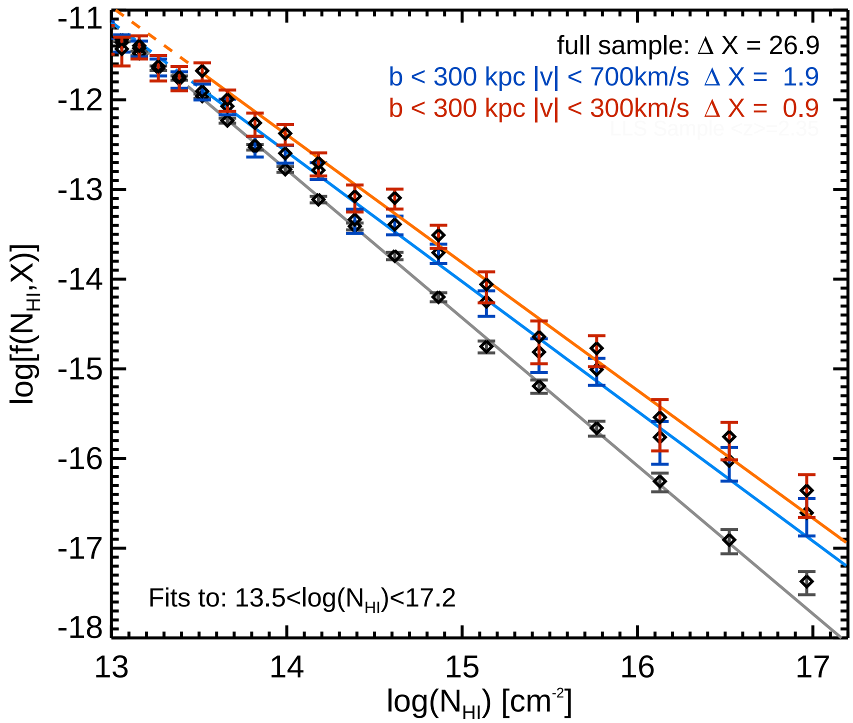

Figure 4. Same as Figure 2 but also considering the distribution of absorbers close to galaxies. The IGM distribution is shown in grey. Red points show the frequency of absorbers of a given $N_{\rm HI}$ within 300 km s^{-1} and 300 pkpc of a galaxy in our spectroscopic sample. Blue points show the frequency of those within 700 km s^{-1} and 300 pkpc. The curves are the maximum likelihood fits to the distributions over the range $13.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$. The parameters of the fit are given in Table 4 along with the opacities calculated from the data and the opacities implied by the fit. Notably this single power law fit to the data does not well reproduce the opacities calculated from the data. Section 5.1 discusses broken-power law fits which better reproduce these opacities.

Using the same KBSS galaxy sample as in Rudie et al. (2012), we measure the frequency distribution of H I absorbers within the CGM as follows. First we locate those galaxies in our spectroscopic sample lying within 300 pkpc of the line of sight to the QSO and within the redshift range used for the Lyα forest decompositions (as listed in Table 1). Notably, we have no galaxies at impact parameters < 50 pkpc, and so the CGM distributions apply to the range $50-300$ pkpc. We then consider the column density distribution of absorbers within ±300 km s^{-1} and ±700 km s^{-1} of the redshifts of these galaxies. We note that all absorbers included in the 300 km s^{-1} CGM are also included in the 700 km s^{-1} CGM.

The resulting frequency distributions for the 300 km s^{-1} and 700 km s^{-1} CGM are shown in Figure 4 in red and blue respectively. As found by Rudie et al. (2012), the frequency of all absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 13.5$ is higher within the CGM of galaxies than at random places in the IGM.

The maximum likelihood estimates of the power law approximation to $f(N_{\rm HI}, X)$ within the CGM are listed in Table 4 and overplotted in Figure 4. Notably, the power law slope of the CGM is significantly shallower than that of the IGM, in the sense that there are relatively more high-$N_{\rm HI}$ systems than low-$N_{\rm HI}$ absorbers in the CGM. This result is qualitatively equivalent to results presented by Rudie et al. (2012) who found a strong correlation between $N_{\rm HI}$ and the fraction of absorbers found within the CGM of galaxies (their Figure 30). Further, the frequency of absorbers is higher within the 300 km s^{-1} CGM than within 700 km s^{-1}, as expected.[7]

The characteristic $N_{\rm HI}$ of the CGM is akin to the max($N_{\rm HI}$) statistic presented in the left-hand panel of Figure 9 of Rudie et al. (2012). These authors typically found $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 14.5-15$ within 300 pkpc of a galaxy. Misawa et al. (2007) considered the frequency distribution of absorbers surrounding H I systems with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 15$, fitting to the absorber distribution within ±1000 km s^{-1}of the redshifts of these tracer absorbers. Thus, the Misawa et al. (2007) sample traces a volume similar to that of the CGM. For their S2a sample which was selected to lie at least 5000 km s^{-1} lower than the redshifts of the QSOs, Misawa et al. (2007) find a power-law slope of $\beta = 1.390 \pm 0.027$, in reasonable agreement with the power-law exponents for our CGM measurements. The normalization, $C_{\rm HI} = 7.262 \pm 0.431$, is lower than our CGM results, consistent with the expectation that the Misawa et al. (2007) sample includes a larger fraction of "general" IGM compared to CGM.

5. THE LYMAN CONTINUUM OPACITY

One of the principal uses of the column density distribution presented above is to estimate the transmissivity of the IGM

[7] Since the majority of the excess absorption is found within 300 km s^{-1} of galaxies, the 700 km s^{-1} CGM is diluted by regions with H I properties more similar to the general IGM. As such, the difference between the 300 and 700 km s^{-1} CGM is mostly a result of the difference in pathlength considered. The pathlength for the 300 km s^{-1} CGM sample is ΔX = 0.9 while the 700 km s^{-1} sample has twice the pathlength, ΔX = 1.9 - and since $f(N_{\rm HI}, X)$ is inversely proportional to the pathlength, the frequency of absorbers in the 700 km s^{-1} CGM is approximately a factor of 2 smaller.

Table 4
MLE Power Law fits to $f(N_{\rm HI}, X)$

Sample	β	$\log(C_{\rm HI})$	$\kappa_{\rm fit}$ [pMpc^{-1}]	$\kappa_{\rm data}$[a] [pMpc^{-1}]
IGM	1.650 ± 0.017	10.322	0.0029	$0.0039^{+0.0005}_{-0.0005}$
CGM 700km s^{-1} 300 pkpc	1.447 ± 0.033	7.672	0.013	$0.017^{+0.004}_{-0.004}$
CGM 300km s^{-1} 300 pkpc	1.425 ± 0.037	7.562	0.022	$0.031^{+0.009}_{-0.009}$

[a] Opacity calculated from the the data using equation 15 for $13.5 \leq \log(N_{\rm HI}/{\rm cm}^{-2}) \leq 17.2$.

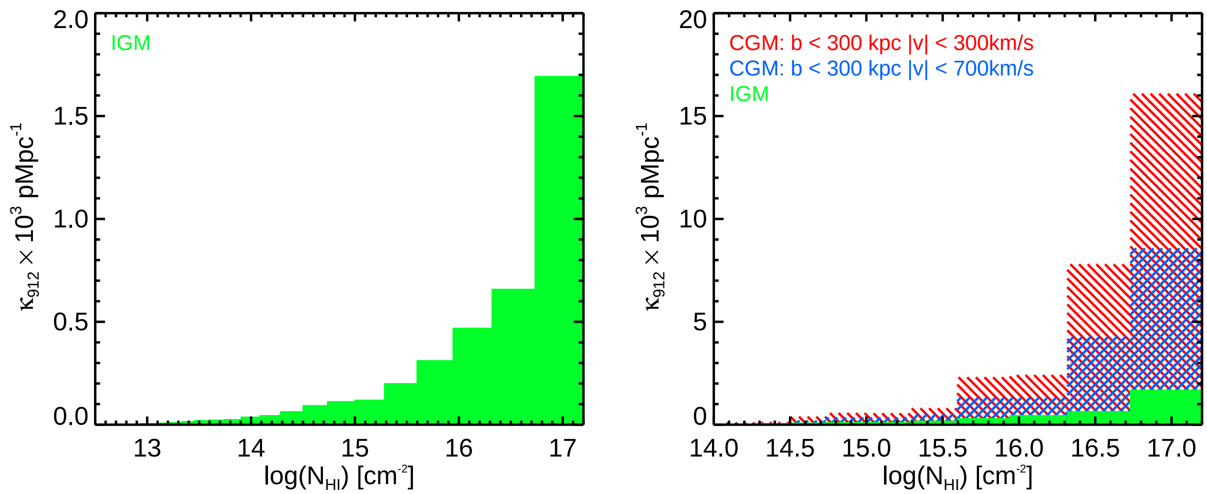

Figure 5. Contributions to the opacity from a given bin in $N_{\rm HI}$, measured directly from the data using equation 15. *Left:* The IGM opacity per bin in $N_{\rm HI}$. *Right:* The opacity per bin in $N_{\rm HI}$ within the IGM (green), the 700 km s^{-1} CGM (blue) and the 300 km s^{-1} CGM (red). Note that the scale of the left-hand panel is 10% of the scale of the right-hand panel. The opacity (optical depth per unit physical distance) of the CGM is an order of magnitude larger than the average opacity in the IGM.

to hydrogen-ionizing photons, hereafter referred to as Lyman Continuum (LyC) photons. Understanding LyC emission and its transmission through the IGM is an important step in exploring the process of reionization at high-redshift, as well as estimating the average ionization level of the IGM and the metagalactic ionizing background as a function of redshift.

In order to understand and describe the process of transmission of LyC photons through the IGM, one must first quantify the sources of opacity in the IGM. Most notably, it is important to quantify the fractional contribution of absorbers of various $N_{\rm HI}$ to the full opacity of the forest. To do this, we begin with a simplified model of LyC emission and absorption.

The attenuation of LyC photons by an absorber with $N_{\rm HI}$= N_i is given by $e^{-N_i\sigma}$ where σ is the photoionization cross section of hydrogen. In practice, σ depends sensitively on the energy of the photons considered:

$$\sigma(\nu) \approx \sigma_{\rm LL} \left(\frac{\nu}{\nu_{\rm LL}}\right)^{-3} \tag{11}$$

where $\sigma_{\rm LL} = 6.35 \times 10^{-18}$ cm^2 is the hydrogen ionization cross section to photons at the Lyman Limit (912 Å) with energies of 1 Ry (13.6 eV), and $\nu_{\rm LL}$ is the frequency of such photons. For the moment, we will employ the simplifying assumption that all LyC photons have energies of 1 Ry so that the frequency dependence of σ is removed.

In addition, for the time being, we calculate the opacity assuming that the universe is static.[8] In practice, the expansion of the universe results in the redshifting of ionizing photons to non-ionizing energies; however since we are only considering 1 Ry photons, we will also temporarily ignore redshifting. In §6, we perform a more rigorous calculation to measure the true mean free path of LyC photons through the IGM and CGM at z=2.4 including photons of all ionizing energies and including redshifting. The analytic calculation presented in this section is provided to build intuition regarding the sources of opacity in the universe, and also to quantify the opacity from absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ inferred directly from the data.

Using the above approximations, we calculate the opacity, κ, of the IGM and CGM to LyC photons and the contribution to this opacity from absorbers of various $N_{\rm HI}$. The opacity is defined as the change in optical depth per unit proper distance:

$$\kappa = \frac{d\tau}{dr} \tag{12}$$

where

$$dr = \frac{1}{1+z}\frac{c}{H(z)}dz \tag{13}$$

and c is the speed of light.

[8] For the remainder of Section 5 we use the redshift distribution of the sample to infer the physical distance over which absorbers are distributed. This is the only way in which redshifts enter into the calculation.

Given an analytic fit to the frequency distribution per unit physical distance $f(N_{\rm HI}, r)$, the opacity due to Poisson-distributed absorbers with $N_{\rm min} \leq N_{\rm HI} \leq N_{\rm max}$ to photons of energy 13.6 eV is:

$$\kappa_{\rm fit} = \int_{N_{\rm min}}^{N_{\rm max}} f(N_{\rm HI}, r) \left(1 - e^{-N_{\rm HI}\sigma_{912}}\right) dN_{\rm HI} \quad (14)$$

(Paresce et al. 1980).

Similarly, the opacity can be calculated directly from the data by summing over those absorbers {i} with $N_{\rm min} < N_{\rm HI} < N_{\rm max}$:

$$\kappa_{\rm data} = \frac{\sum_i \left(1 - e^{-N_{\rm HI,i}\sigma_{912}}\right)}{\Delta r} \quad (15)$$

where Δr is the sum of the proper distances probed by each sightline.

The uncertainty in the measurement of $\kappa_{\rm data}$ can be calculated using the bootstrap resampling method in which many sets of absorbers are drawn with replacement and $\kappa_{\rm data}$ is recalculated for each set. One can then estimate the probability distribution of $\kappa_{\rm data}$ by selecting the 68% confidence range. Values of $\kappa_{\rm data}$ for the IGM and CGM for $13.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ is listed in the last column of Table 4. *We emphasize that this measurement of the LyC opacity from absorbers with* $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17$ *is one of the principal results of this work.*

The left hand panel of Figure 5 shows the opacity produced by various bins of $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ within the IGM calculated directly from the data using equation 15. These values are compared to the opacity found within the CGM in the right-hand panel of Figure 5. Notably, the opacity of the CGM of a galaxy (300 km s^{-1}, 300 pkpc) is an order of magnitude larger than that of the average IGM.

Comparing the opacities calculated using the single power law MLE fit with those calculated directly from the data for absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ (Table 4), one notes a statistically significant discrepancy in the sense that the MLE power law predicts far less opacity than is measured directly from the data. Similarly, the number of LLSs (which also contribute significantly to the LyC opacity) is under-predicted by the MLE fit. In Section 5.1, we consider other parameterizations of the frequency distribution that better match $dn_{\rm LLS}/dX$ as well as the opacity from intermediate-$N_{\rm HI}$ systems [$15 \lesssim \log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 17$] calculated from the data.

5.1. *Broken Power Law Parameterizations*

While the MLE power law fits presented in §3.2 and §4 provide a reasonable description of the overall column density distribution, because there are substantially more absorbers at low-$N_{\rm HI}$ the majority of the statistical power for the fit is derived from the low-$N_{\rm HI}$ absorbers. To estimate the LyC opacity of the intergalactic and circumgalactic gas, the absorbers of relevance are instead the much rarer, high-$N_{\rm HI}$ systems, particularly those near $\log(N_{\rm HI}/{\rm cm}^{-2}) \approx 17$.

In order to reproduce simultaneously the opacity from both absorbers with $12.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ as measured in our sample, as well as the number of LLSs, we adopt a different fitting method. Here we consider a least-squares fit to the binned[9] data. Because the opacity is a strong function of $N_{\rm HI}$, it is most important that the number of absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) \approx 17.2$ is accurately reproduced. Therefore, to provide further flexibility, we allow for a discontinuous break in the power law parameterization. This allows for an independent fit to the high-$N_{\rm HI}$ absorbers [$\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 15$] which are the main source of opacity. Although no strong break is evident, any parameterization that does not include a break as well as all parameterizations with a continuous functional form at the position of the break cannot reproduce the opacity of Lyα forest absorbers as measured by the data.

We fit two power laws to the binned data. The bin that includes the break is included in the fit to the low and intermediate-$N_{\rm HI}$ groups to minimize the discontinuity between the two power laws. We allow the break to vary in $N_{\rm HI}$ and select the break location to be that which best reproduces the opacity from absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ and also the number of LLS in the case of the IGM sample. The break location for the two CGM distributions is forced to match. The best fits are presented in Figure 6 and Table 5 along with their implied opacities, $\kappa_{\rm fit}$. The bin containing the break is indicated by the hatched box in Figure 6. Notably, the location of the break ($N_{\rm HI} \approx 10^{15}$ cm^{-2}) is similar to the $N_{\rm HI}$ at which absorbers were found to significantly correlate with the positions of galaxies (Rudie et al. 2012).

This fitting technique reproduces the opacity from intermediate-$N_{\rm HI}$ absorbers measured from the data with high fidelity. Additionally, the IGM fit predicts a $dn_{\rm LLS,fit}/dX = 0.519$ for $17.2 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 22.0$ in excellent agreement with $dn_{\rm LLS,data}/dX = 0.52 \pm 0.08$ determined from the compilation of LLSs from the literature (Section 3.3).[10] As in section 3.3, the green triangle in Figure 6 represents the values of $f(N, X)$ within the LLS regime implied by the measurement of $dn_{\rm LLS,data}/dX$ *assuming a power law index*, $\beta = \beta_{\rm high-N_{HI}}$. The agreement between the height on the green triangle and the extrapolation of the power law fit (grey line) suggests that the broken power law parameterization well reproduces the LLS gross statistics.

Next we consider the uncertainty in the fitted parameters. The majority of the uncertainty in the normalization of the frequency distribution ($C_{\rm N_{HI}}$) is due to its covariance with the power law index, β. In Table 6 we list the best fit determination of $\beta_{\rm high-N_{HI}}$ and $C_{\rm high-N_{HI}}$ and the values of $\kappa_{\rm fit}$ and $dn_{\rm LLS,fit}/dX$ implied by this analytic approximation. Also listed are the 1-σ perturbed values of $\beta_{\rm high-N_{HI}}$, the corresponding normalization $C_{\rm high-N_{HI}}$, and their implied $\kappa_{\rm fit}$ and $dn_{\rm LLS,fit}/dX$. Notably, the perturbed analytic parameters reproduce the opacity due to Lyα forest absorbers within the measurement errors. The incidence of LLSs is more sensitive to perturbed values of β with offsets in the implied incidence of $\sim 1.5\sigma$ from the observed incidence. As the incidence of LLSs is not used to constrain the power-law fit, the disagreement between these values is not surprising.

Figure 7 compares the IGM broken power law fit to the frequency of LLS with $\tau_{912} > 1$ and $\tau_{912} > 2$ from O'Meara et al. (2012) as described in §3.3. The blue triangle corresponds to $\tau_{912} > 1$ absorbers, and the red triangle represents the $\tau_{912} > 2$ absorbers. As in section 3.3, the normalization of the triangles are set by the measured value of $dn_{\rm LLS,data}/dX$ *under the assumption that an extrapolation of the broken power law fit to the Lyα forest data holds within the LLS regime.* Again, the agreement between the extrapolated power-law fit (grey

[9] We have verified that the fitted parameters are insensitive to the exact choice of binning.

[10] We remind the reader that the single power law MLE determination of the fit predicted a value of $dn_{\rm LLS,fit}/dX = 0.21$.

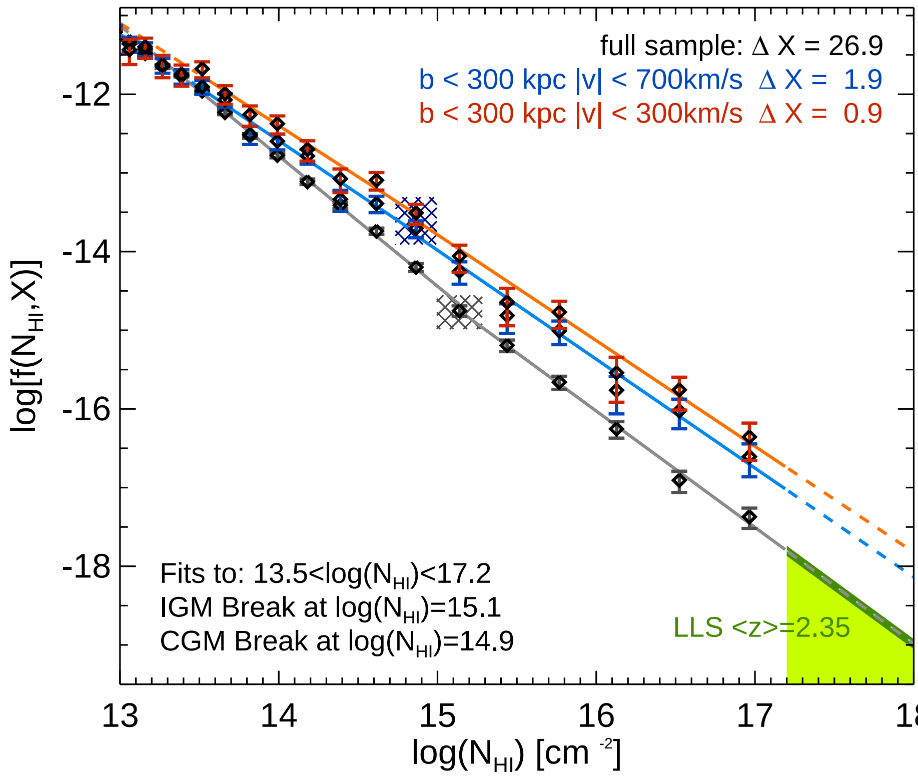

Figure 6. Broken power law least-squares fits to the binned IGM (grey) and CGM (blue and red) data. The slopes and normalizations for each data set (IGM or CGM) are allowed to vary independently; however, both ranges include the data point centered at the break (shaded regions). Note that the broken power-law fit is discontinuous at the break location. Including one data point in both ranges minimizes the discontinuity associated with the break and produces a better fit to the data. The hatched box shows the range of $N_{\rm HI}$ included in that point and thus in both fits. The green triangle represents the measured incidence of LLSs as discussed in §3.3. The parameters of the fit are given in Table 5 along with the opacity calculated from the data and the opacity produced by the fit to the distribution. Note that the opacity is reproduced to much higher fidelity using the broken power law fit.

Table 5
Broken Power Law fits to $f(N_{\rm HI}, X)$

Sample	$\beta_{\rm low-N_{HI}}$	$\log(C_{\rm low-N_{HI}})$	break $N_{\rm HI}$	$\beta_{\rm high-N_{HI}}$	$\log(C_{\rm high-N_{HI}})$	$\kappa_{\rm fit}$ [pMpc^{-1}]	$\kappa_{\rm data}$[a] [pMpc^{-1}]
IGM:	1.656± 0.030	10.398	15.14	1.479± 0.042	7.643	0.0038	$0.0039^{+0.0005}_{-0.0005}$
CGM 700km s^{-1} 300 pkpc	1.353± 0.088	6.354	14.87	1.385± 0.084	6.789	0.017	$0.017^{+0.004}_{-0.004}$
CGM 300km s^{-1} 300 pkpc	1.302± 0.084	5.823	14.87	1.345± 0.089	6.394	0.031	$0.031^{+0.009}_{-0.009}$

[a] Opacity calculated from the the data using equation 15 for $13.5 \le \log(N_{\rm HI}/{\rm cm}^{-2}) \le 17.2$.

curve) and the blue and red triangles suggests that the analytic approximation to the frequency distribution reproduces the measure LLS statistics from O'Meara et al. (2012). The green point in Figure 7 shows the measured frequency of absorbers with $1 < \tau_{912} < 2$ [$17.2 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.5$] calculated from the difference between the measured values of $dn_{{\rm LLS},\tau>1}/dX$ and $dn_{{\rm LLS},\tau>2}/dX$.

The power law fit to $f(N,X)$ suggested by O'Meara et al. (2012) is overplotted as the red dotted line for comparison. This curve would imply a value for the opacity from absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ of $\kappa_{\rm fit} = 0.0022$, nearly a factor of two smaller than that measured from the KBSS sample ($\kappa_{\rm data} = 0.0039 \pm 0.0005$).

There are likely some deviations in the detailed distribution of $f(N_{\rm HI}, X)$ in the subDLA ($19.3 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 20.3$) and DLA ($\log(N_{\rm HI}/{\rm cm}^{-2}) > 20.3$) regimes as discussed in Prochaska et al. (2009) and O'Meara et al. (2012). We show the current constraints on these regions of the frequency distribution in the right hand panel of Figure 7. We chose to neglect the differences between the power law fit and observed constraints on $f(N,X)$. Holding the measured value of $dn_{\rm LLS}/dX$ fixed, such an assumption makes our measurements of the opacity from absorbers with $N_{\rm HI} > 10^{17.2}$ cm^{-2} a *lower limit*. However, since the attenuation resulting from any system with $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 18$ is similar, the detailed shape of the distribution for high-$N_{\rm HI}$ will introduce relatively small changes in the opacity so long as the total number of such systems is reproduced.

Table 6
Effect of uncertainty in β on κ and $dn_{\rm LLS}/dX$[a]

Sample	$\beta_{\rm high-N_{HI}}$	$\log(C_{\rm high-N_{HI}})$[b]	κ [pMpc^{-1}][c]	$dn_{\rm LLS}/dX$[d]
Data[e]	···	···	$0.0039^{+0.0005}_{-0.0005}$	0.52 ± 0.08
IGM Best Fit	1.479	7.643	0.0038	0.519
IGM Shallow Fit[f]	1.437	6.984	0.0040	0.656
IGM Steep Fit[g]	1.521	8.301	0.0035	0.413

[a] The effect of varying $\beta_{\rm high-N_{HI}}$ within its 1-σ uncertainty.
[b] The best fit normalization of $f(N,X)$ given the stated value of β. I.e., the co-variance of $\log(C_{\rm high-N_{HI}})$ with $\beta_{\rm high-N_{HI}}$.
[c] The opacity from absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$.
[d] The incidence of LLSs
[e] Values computed directly from the data.
[f] The fit for the perturbed value of $\beta_{\rm high-N_{HI}}$ such that the distribution is more shallow.
[g] The fit for the perturbed value of $\beta_{\rm high-N_{HI}}$ such that the distribution is steeper.

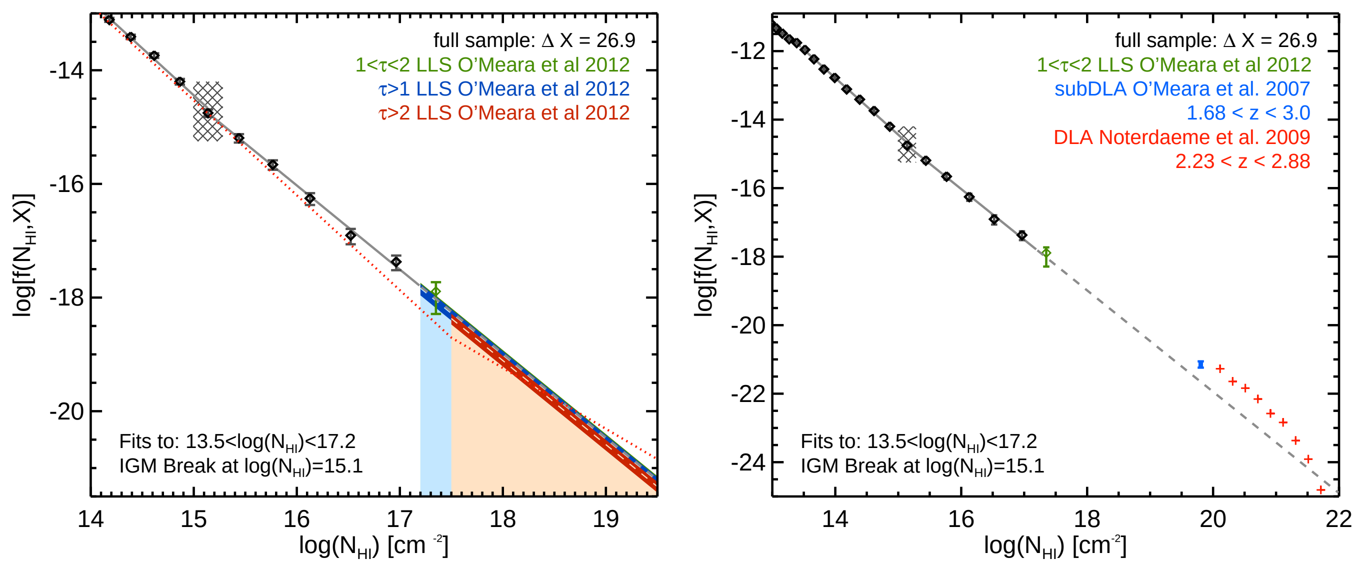

Figure 7. Same at Figure 6 but only considering the IGM distribution. *Left:* The results of a recent survey for Lyman limit absorbers by O'Meara et al. (2012). Note that these authors' constraints on $\tau_{912} > 1$ and $\tau_{912} > 2$ LLS are well reproduced by the extrapolation of the power law slope selected for the high-$N_{\rm HI}$ IGM absorbers. Computing the difference between these values allows for an assessment of the number of LLSs with $1 < \tau_{912} < 2$, here represented as the green data point. The (red) dotted curve shows the fit to the frequency distribution proposed by O'Meara et al. (2012). Note that this fit underpredicts the number of absorbers with $\log(N_{\rm HI})/{\rm cm}^{-2}) < 17.2$ as measured with the KBSS sample (black data points). *Right:* Including all currently known constraints at these redshifts for $f(N,X)$. In green the $1 < \tau_{912} < 2$ measurement from O'Meara et al. (2012). In blue are the subDLAs ($19.3 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 20.3$; $dn_{\rm subDLA}/dX = 0.110^{+0.026}_{-0.021}$) from O'Meara et al. (2007) and in red are the DLAs ($\log(N_{\rm HI}/{\rm cm}^{-2}) > 20.3$; $dn_{\rm DLA}/dX \approx 0.05$) from Noterdaeme et al. (2009). We note that the extrapolation of the intermediate-$N_{\rm HI}$ absorbers power law fit suggested by our Lyα forest data appears to provide a good fit to the data for $\log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 18$. Above this $N_{\rm HI}$, nature likely deviates significantly from this power law.

The opacity inferred from the power law fit to $f(N,X)$ is a lower limit for the following reason. If we hold $dn_{\rm LLS}/dX$ fixed at the observed value, this gives us a constraint on the number of systems with $N_{\rm HI} > 10^{17.2}$ cm^{-2}. The assumption of a given value of the power-law index, β, allows us to infer the opacity from a measurement of $dn_{\rm LLS}/dX$. Allowing β to take a shallower value to match the plotted points of $f(N,X)$ for the subDLAs and DLAs only moves absorbers to higher $N_{\rm HI}$ and thus higher opacity compared to a model with a steeper power-law index. Therefore, our suggested opacities are likely a slight under-prediction of the true value. If we instead choose a flat value ($\beta = 0$) we would find 9.7% higher opacity from LLSs, suggesting the total uncertainty in the opacity from the assumed value of β is significantly less than 10%.

5.2. *The Opacity of the IGM and CGM*

Under the assumptions outlined above, and with a parameterization of the frequency distribution that reproduces both the incidence of LLS and the opacity from intermediate-$N_{\rm HI}$ absorbers, we can now measure the full opacity of the IGM and CGM. We calculate the opacity from LLS absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$ using equation 14[11] and the opacity

[11] Note that the detailed distribution of κ per bin in $N_{\rm HI}$ would be affected by variation in $f(N_{\rm HI})$ at DLA and subDLA column densities. Because we

Table 7
Opacity and Optical Depth of Absorbers of various $N_{\rm HI}$

$N_{\rm HI}$ min	$N_{\rm HI}$ max	method	$\kappa_{\rm IGM}$ $\times 10^3$ [pMpc^{-1}]	$\tau_{\rm IGM}$ $\times 100$	$\kappa_{\rm CGM\ 700}$ $\times 10^3$ [pMpc^{-1}]	$\tau_{\rm CGM\ 700}$ $\times 100$	$\kappa_{\rm IGM}$ $\times 10^3$ [pMpc^{-1}]	$\tau_{\rm IGM}$ $\times 100$	$\kappa_{\rm CGM\ 700}$ $\times 10^3$ [pMpc^{-1}]	$\tau_{\rm CGM\ 700}$ $\times 100$
			Differential: $N_{\rm HI}$ min $< \log(N_{\rm HI}/{\rm cm}^{-2}) <$ $N_{\rm HI}$ max				Cumulative: $12.55 < \log(N_{\rm HI}/{\rm cm}^{-2}) <$ $N_{\rm HI}$ max			
12.55	12.60	data	0.001	0.02	0.001	0.00	0.001	0.02	0.001	0.00
12.60	12.65	data	0.002	0.02	0.002	0.00	0.003	0.03	0.002	0.00
12.65	12.71	data	0.002	0.02	0.002	0.00	0.005	0.06	0.004	0.00
12.71	12.77	data	0.003	0.03	0.003	0.00	0.008	0.09	0.007	0.00
12.77	12.85	data	0.004	0.04	0.003	0.00	0.012	0.13	0.009	0.00
12.85	12.93	data	0.004	0.04	0.003	0.00	0.016	0.17	0.012	0.00
12.93	13.01	data	0.006	0.07	0.006	0.00	0.022	0.24	0.018	0.01
13.01	13.11	data	0.008	0.08	0.007	0.00	0.030	0.32	0.024	0.01
13.11	13.21	data	0.009	0.10	0.010	0.00	0.039	0.42	0.034	0.01
13.21	13.32	data	0.011	0.12	0.011	0.00	0.050	0.54	0.046	0.01
13.32	13.45	data	0.017	0.18	0.016	0.00	0.067	0.72	0.061	0.02
13.45	13.59	data	0.021	0.22	0.022	0.01	0.088	0.94	0.083	0.02
13.59	13.74	data	0.024	0.25	0.032	0.01	0.112	1.20	0.116	0.03
13.74	13.90	data	0.027	0.29	0.027	0.01	0.140	1.49	0.143	0.04
13.90	14.08	data	0.038	0.40	0.053	0.02	0.178	1.89	0.195	0.06
14.08	14.28	data	0.045	0.48	0.089	0.03	0.223	2.37	0.285	0.08
14.28	14.50	data	0.066	0.70	0.073	0.02	0.289	3.07	0.358	0.10
14.50	14.73	data	0.095	1.01	0.198	0.06	0.383	4.08	0.557	0.16
14.73	15.00	data	0.115	1.23	0.353	0.10	0.499	5.31	0.910	0.26
15.00	15.28	data	0.123	1.30	0.351	0.10	0.621	6.61	1.261	0.36
15.28	15.60	data	0.202	2.14	0.511	0.15	0.823	8.75	1.772	0.51
15.60	15.94	data	0.314	3.34	1.320	0.38	1.136	12.09	3.092	0.89
15.94	16.32	data	0.471	5.01	1.316	0.38	1.607	17.10	4.408	1.26
16.32	16.73	data	0.661	7.03	4.278	1.23	2.268	24.13	8.686	2.49
16.73	17.20	data	1.695	18.03	8.571	2.46	3.963	42.17	17.257	4.95
17.20	17.70	fit	1.570	16.70	9.249	2.65	5.534	58.87	26.506	7.60
17.70	18.20	fit	1.111	11.82	7.273	2.09	6.645	70.69	33.779	9.68
18.20	18.70	fit	0.647	6.88	4.719	1.35	7.291	77.57	38.498	11.04
18.70	19.20	fit	0.372	3.96	3.030	0.87	7.664	81.53	41.527	11.91
19.20	19.70	fit	0.214	2.28	1.946	0.56	7.878	83.81	43.473	12.46
19.70	20.30	fit	0.141	1.50	1.439	0.41	8.019	85.31	44.912	12.88
20.30	22.00	fit	0.127	1.35	1.597	0.46	8.146	86.67	46.509	13.33

[a] Opacity calculated from the the data using equation 15 for $12.55 \le \log(N_{\rm HI}/{\rm cm}^{-2}) \le 17.2$.

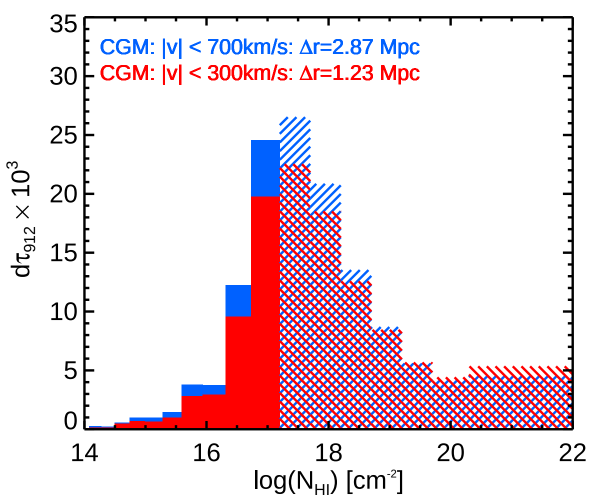

Figure 8. The optical depth of the CGM to13.6 eV photons. The bins with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ (solid shading) are measured directly from the data. The bins corresponding to LLSs ($\log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$; hatched shadding) use the broken power law fit to the higher $N_{\rm HI}$ end of $f(N_{\rm HI}, X)$ as shown in Figure 6. The blue and red histograms show the total optical depth within one average CGM. This is equivalent to the opacity times the scale length $\Delta r_{\rm CGM}$. Note that the total optical depth in the 700 km s^{-1} CGM includes the opacity within the 300 km s^{-1} CGM but that there exists additional opacity due to absorbers with velocity offsets between ± 300 and ± 700 km s^{-1}.

from intermediate-$N_{\rm HI}$ absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ directly from the data using equation 15 (as described in the beginning of Section 5). The differential opacity from bins in $N_{\rm HI}$ as well as the cumulative opacity (from absorbers with $N_{\rm HI}$ smaller than a given value), are given in Table 7. The total opacity of the IGM is measured to be:

$$\kappa_{\rm IGM} = \kappa_{\rm data}\left(N_{\rm HI} < 10^{17.2}\right) + \kappa_{\rm fit}\left(N_{\rm HI} > 10^{17.2}\right) \quad (16)$$

$$\kappa_{\rm IGM} = 8.146 \times 10^{-3}\ {\rm pMpc}^{-1}. \quad (17)$$

Similarly,

$$\kappa_{\rm CGM\ 300\ km\ s^{-1}} = 94.4 \times 10^{-3}\ {\rm pMpc}^{-1} \quad (18)$$

$$\kappa_{\rm CGM\ 700\ km\ s^{-1}} = 46.5 \times 10^{-3}\ {\rm pMpc}^{-1}. \quad (19)$$

Another relevant quantity is the integrated opacity over a specific scale length, i.e., the optical depth τ. In particular, it is now possible to calculate the optical depth of the CGM of an individual galaxy. This quantity is particularly relevant for understanding the attenuation of ionizing flux emanating from a galaxies similar to those in the KBSS sample due to the CGM of the galaxy itself. Here we attempt to separate the probability that ionizing photons escape the ISM of a galaxy from the probability that photons, having already escaped the ISM, will also escape the enhanced absorption found within the CGM.

adopt the extrapolation of the broken-power law fit to $f(N, X)$ to calculate the opacity, the values in individual bins of $N_{\rm HI}$ for $\log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$ are not well constrained.

Specifically, the KBSS data set constrains the opacity and optical depth of the CGM of galaxies at impact parameters greater than 50 kpc[12]. Henceforth, we consider the classical "escape fraction" to be the probability that an ionizing photon reaches a distance of 50 pkpc, while the CGM is defined to begin at a distance of 50 pkpc from the galaxy. In order to calculate the optical depth, the scale length of interest must be assumed. For circumgalactic absorption, the velocity window associated with the CGM provides a natural scale length. As the relevant scale is along the line of sight, we use the distance, $\Delta r_{\rm CGM}$, associated with the line-of-sight velocity amplitude of the CGM:

$$\Delta r_{\rm CGM} = H^{-1}(z) d\mathrm{v} \tag{20}$$

where $H(z)$ is the Hubble constant evaluated at the mean redshift of the sample $\langle z \rangle = 2.4$ and $d\mathrm{v}$ is either 700 km s^{-1} or 300 km s^{-1}. We note that the KBSS data set constrains the distribution of gas at transverse distances > 50 pkpc from galaxies, not along the line of sight from galaxies. Here we assume that the gas distribution traced in the transverse plane well represents the gas distribution along the line of sight to a galaxy at distances > 50 kpc. Given the strong increase in $N_{\rm HI}$ at small galactocentric distances [as observed in Rudie et al. (2012)], the optical depth associated with the CGM at distances > 50 kpc is likely significantly smaller than the optical depth associated with the closer-in CGM. Therefore the measurements presented here might be thought of as a lower limit on CGM absorption.

In Figure 8, we compare the optical depth of the CGM ($\tau_{\rm CGM} = \kappa_{\rm CGM} \Delta r_{\rm CGM}$) for the 300 km s^{-1}, 300 pkpc CGM (red) and the 700 km s^{-1}, 300 pkpc CGM (blue). Figure 8 also quantifies the optical depths from both forest absorbers ($\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$) and LLS in the CGM. The forest absorber opacities and optical depths are measured directly from the data, while the LLS opacity and optical depths are inferred from an extrapolation of the high-$N_{\rm HI}$ fit to the data using the broken power law parameterization. The values for $\tau_{\rm CGM}$ and $\kappa_{\rm CGM}$ in differential and cumulative $N_{\rm HI}$ bins are given in Table 7.

We find that the optical depth of the 700 km s^{-1} CGM is comparable to but slightly exceeds the optical depth in the 300 km s^{-1} CGM. This is as expected because the 300 km s^{-1} CGM includes the majority of the excess absorption; however, there is a small additional excess between 300 and 700 km s^{-1}, see Rudie et al. (2012). For subsequent calculations, only the 700 km s^{-1} CGM will be considered since it captures the full optical depth of the CGM for $d > 50$ pkpc.

5.3. *The Distribution of Optical Depths in the IGM and CGM*

With an estimate of the opacity from all $N_{\rm HI}$ absorbers, one can easily calculate the fractional contribution of the CGM vs the IGM, as well as the contribution from absorbers of different $N_{\rm HI}$, to the mean free path. For this simplified calculation, we define the mean free path, $\lambda_{\rm mfp}^{\rm static}$, to be the average distance an ionizing photon with energy 1 Ry travels in a static universe before having a probability of $1-e^{-1}$ of being absorbed. More quantitatively:

$$\lambda_{\rm mfp}^{\rm static} = \sum_i \Delta r_i \tag{21}$$

[12] This is because the closest galaxy to a QSO sightline in the KBSS sample is at a projected physical distance of 50 kpc.

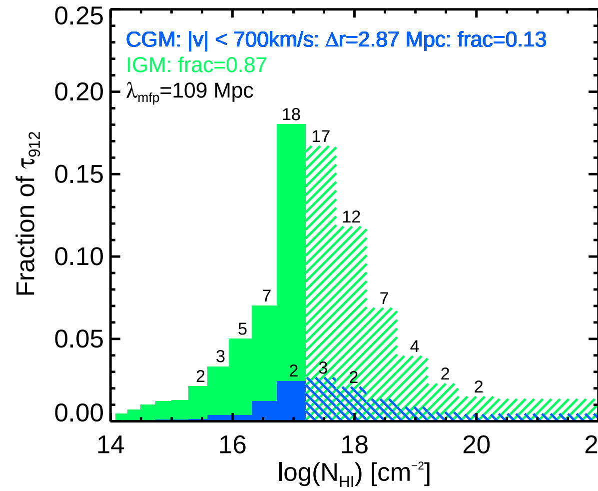

Figure 9. The fraction of the optical depth within 1 mean free path contributed by absorbers within a given bin of $N_{\rm HI}$ in the IGM (green) or 700 km s^{-1} CGM (blue). The solid histograms represent measurements from the data converted into an optical depth using equations 12 and 15. The hatched histograms represent values from the fit to the data extrapolated to higher values of $N_{\rm HI}$ and then converted into an optical depth using equations 12 and 14. The printed numbers give the percentage of the opacity due to that $N_{\rm HI}$ bin.

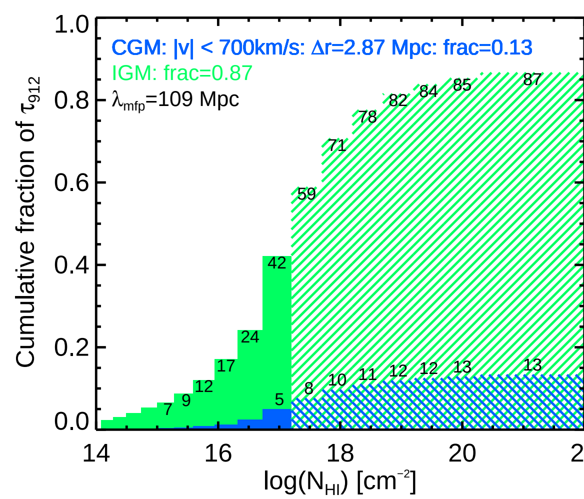

Figure 10. The cumulative fraction of the optical depth within 1 mean free path contributed by absorbers below a given $N_{\rm HI}$ in the IGM (green) or 700 km s^{-1} CGM (blue). The solid histograms represent measurements from the data converted into an optical depth using equations 12 and 15. The hatched histograms represent values from the fit to the data extrapolated to higher values of $N_{\rm HI}$ and then converted into an optical depth using equations 12 and 14. The printed numbers give the percentage of the opacity due to absorbers with $N_{\rm HI}$ less than or within the $N_{\rm HI}$ bin.

where

$$\sum_i \kappa_i \Delta r_i = 1 \tag{22}$$

Here we consider two possible mean free paths: (1) In the case of photons emanating from the ISM of a galaxy similar to those in our spectroscopic sample (see Section 4), they must first traverse the CGM of the galaxy in which they were formed and then the general IGM. Here we refer to this mean

free path as $\lambda^{\rm static}_{\rm mfp,\ IGM+CGM}$.(2) The second case is the more general one and the mean free path that has been considered by previous authors - the mean free path of photons traveling in the IGM, a quantity relevant to photons once they have escaped the CGM of whichever structure in which they were formed. This distance is referred to as $\lambda^{\rm static}_{\rm mfp,\ IGM}$.

In the calculation that follows, we consider the contributions to the opacity from the 700 km s^{-1} CGM as well as the IGM. Under these assumptions:

$$\kappa_{\rm CGM}\Delta r_{\rm CGM} + \kappa_{\rm IGM}(\lambda^{\rm static}_{\rm mfp,\ IGM+CGM} - \Delta r_{\rm CGM}) = 1. \qquad (23)$$

where $\Delta r_{\rm CGM}$ = 2.87 pMpc as given in equation 20 using $dv = 700$ km s^{-1}, and $\kappa_{\rm IGM} = 0.0081$ pMpc^{-1} and $\kappa_{\rm CGM} = 0.047$ pMpc^{-1} are the total opacities over all $N_{\rm HI}$. Solving equation 23, we find $\lambda^{\rm static}_{\rm mfp,\ IGM+CGM}$= 109.3 pMpc. When we exclude the CGM component, we instead obtain $\lambda^{\rm static}_{\rm mfp,\ IGM}$= 122.8 pMpc, a 12% longer pathlength.

With $\lambda^{\rm static}_{\rm mfp,\ IGM+CGM}$ known, we can then calculate the fractional contribution to the optical depth from each bin of $N_{\rm HI}$ for both the IGM and CGM as shown in Figure 9 and Table 7 (middle columns). The cumulative version of this plot (i.e. the contribution to the optical depth for all $N_{\rm HI} < N_i$) is given in Figure 10 and in Table 7, right-hand columns. From these, we can see that the CGM accounts for 13% of the opacity within a $\lambda^{\rm static}_{\rm mfp,\ IGM+CGM}$. Further, absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ contribute 48% of the opacity to LyC photons.

In practice, the detailed distribution of opacity as a function of $N_{\rm HI}$ will be altered when photons of various energies are considered. The true $\lambda_{\rm mfp}$ of LyC photons will be non-negligibly affected by the redshifting of photons to lower energies, and by the presence of higher-energy radiation. For instance, 100 pMpc corresponds to $\Delta z \approx 0.25$ at these redshifts. Photons emitted by a source at $\langle z \rangle = 2.4$ with energies close to the Lyman Limit therefore will redshift to non-ionizing energies long before they are attenuated by e.

In the next section, we consider a more rigorous calculation of $\lambda_{\rm mfp}$ including the aforementioned effects using Monte Carlo simulations of discrete lines of sight. As we will show, a more proper treatment of the physics does not substantially affect the value of $\lambda_{\rm mfp}$.

6. THE MEAN FREE PATH TO HYDROGEN-IONIZING PHOTONS

As we have shown in Section 5, the mean free path of LyC photons is a cosmological distance at $z = 2.4$. As such, the consideration of photons with energies > 1 Ry as well as inclusion of redshifting of photons out of ionizing frequencies are necessary in order to measure the true $\lambda_{\rm mfp}$. It is also necessary to define very precisely what one means by the 'mean free path,' including the frame of reference in which the distance is measured. Here we define the mean free path, $\lambda_{\rm mfp}$, as the distance traveled from a source by a packet of photons before photons with energy of 1 Ry in the *absorber's*[13] frame are attenuated by an average factor of e. This distance is fundamentally related to a measurement of Δz, which we convert to the physical distance traveled by the photon through the expanding universe.

We perform a Monte Carlo (MC) simulation of the LyC absorption along many individual sightlines. For illustration, we begin with a discussion of MC runs with an emission redshift of $z = 2.4$.

[13] We consider the energy of the photon at the absorber's redshift because only photons with energy sufficient to ionize hydrogen when they intersect an absorber are relevant to $\lambda_{\rm mfp}$.

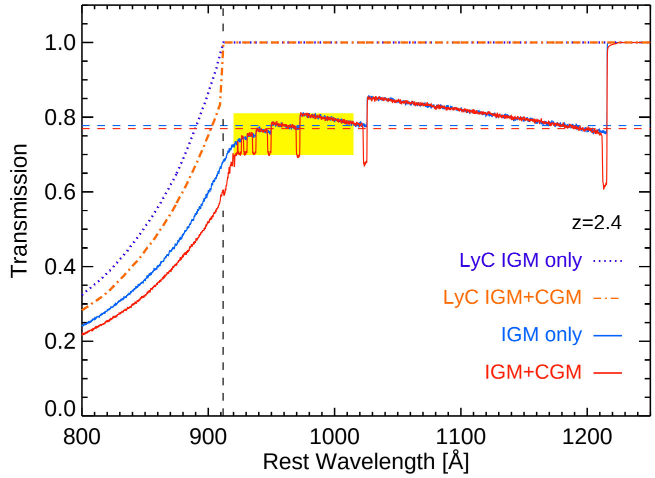

Figure 11. The average transmission spectra from 10,000 runs of the Monte Carlo simulations of IGM only (purple and blue) and IGM+CGM(red and orange) absorption in the foreground of a $z = 2.4$ emitter shown in the rest frame of the emitter. The solid (blue and red) curves show spectra using MC simulations that include line opacity. The dotted (purple) and dot-dashed (orange) curves show spectra from simulations only including continuum opacity. The highlighted yellow area marks the region between 920 -1015 Å in the frame of the emitter which is used to renormalize the line opacity spectra in Figure 12. The dashed horizontal lines mark the average value of the transmission in this region, $\langle 1 - D_B \rangle$, for the spectra containing line opacity. The vertical dashed line marks the position of the Lyman limit in the frame of the emitter. The line absorption seen in the red spectrum is due to the Lyman series line opacity from high-$N_{\rm HI}$ CGM absorbers.

6.1. *Monte Carlo Simulations*

To simulate absorption within intergalactic and circumgalactic space, we produced artificial absorption spectra whose absorption line distribution is matched to that observed in data at the same redshifts; such models are described in detail by Shapley et al. (2006). In this case, we produced 10,000 simulated lines of sight to sources with $z_{\rm em}$ = 2.4, according to a frequency distribution of absorbers set using best-fit parameters from the broken power law fit to $f(N,X)$ (Table 5).

As was demonstrated in Section 5.1, various fitting methods produce largely discrepant values for the parameters of the analytic form of $f(N,X)$. The actual opacity represented by the data can be measured to significantly higher precision then the parameters of the analytic form of $f(N,X)$. For this reason, using perturbed values of the power law parameters will result in significantly larger variation in the opacity along a line of sight than occur in the actual universe as measured by the data. As such, we use only the best-fit values of the power law fit to $f(N,X)$ in the Monte Carlo.

The simulated absorbers are distributed throughout the redshift range of interest in such a way that they reproduce the overall redshift distribution of measured absorbers. The redshift distribution of absorbers is typically parameterized as:

$$\frac{dn}{dz} = n_0(1+z)^{\gamma}. \qquad (24)$$

For the MC simulations, we have assumed that the Lyman-α forest absorbers with low-$N_{\rm HI}$ (those absorbers fit by the

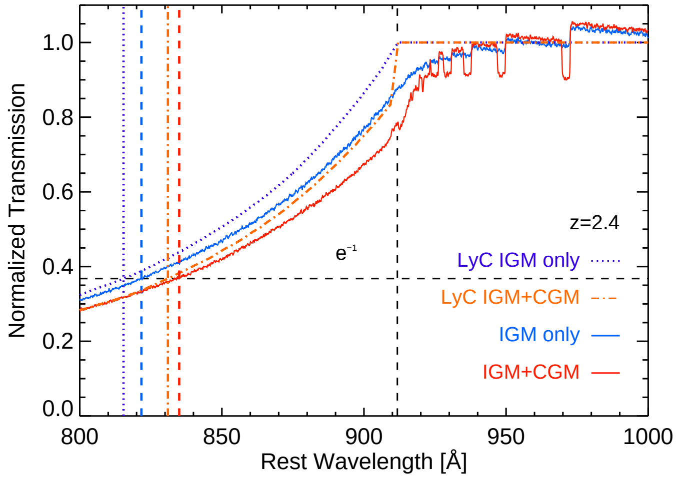

Figure 12. The average normalized transmission below the Lyman limit of a $z = 2.4$ emitter. The red and orange curves pertain to transmission including the CGM and the blue and purple curves pertain to IGM only transmissions. The dotted (purple) and dot-dashed (orange) lines show the same spectra as in Figure 11 from the continuum opacity only runs of the MC code. The solid (red and blue) curves show the spectra including line opacity normalized by $\langle 1-D_B \rangle$. The black vertical dashed line marks the position of the Lyman limit in the frame of the emitter. The dashed horizontal line shows where the normalized transmission is equal to e^{-1}. The colored vertical lines mark the position where each of the spectra crosses the e^{-1}, signifying the position of one $\lambda_{\rm mfp}$. These rest wavelengths are converted into physical distances corresponding to the $\lambda_{\rm mfp}$ as described in the text.

low-$N_{\rm HI}$ power law in the broken power law parameterization: $\log(N_{\rm HI}/{\rm cm}^{-2}) < 15.1$ for the IGM and $\log(N_{\rm HI}/{\rm cm}^{-2}) < 14.9$ for the CGM) evolve with $\gamma = 2.5$ (Kim et al. 2002). We have verified that the line sample presented in this paper exhibits evolution consistent with these measurements. All high-$N_{\rm HI}$ absorbers are assumed to have $\gamma = 1.0$; the value for the LLS compilation reported in this paper (Section 3.3) measured through maximum likelihood estimation is $\gamma = 0.93 \pm 0.29$.

Briefly, we note that the redshift evolution for LLSs within the range $2 \lesssim z \lesssim 3$ is well constrained by the data presented by Ribaudo et al. (2011), Sargent et al. (1989), and Stengler-Larrea et al. (1995) as discussed in Section 3.3. Measurements of the evolution in the line density at $z \gtrsim 3$ (see e.g., Songaila & Cowie 2010; Prochaska et al. 2010) suggest much steeper evolution with redshift (high values of γ). Because our Monte Carlo simulations only include absorption at $2 \lesssim z \lesssim 3$, we use the value of $\gamma = 1$ as measured over the relevant redshift range and emphasize that the exact value of γ does not significantly affect the quoted $\lambda_{\rm mfp}$ within the redshift range $2 < z < 3$. [14] However, we caution that the extrapolations of the trend of our $\lambda_{\rm mfp}$ measurements to $z > 3$ is much more dependent on the chosen value of γ and is therefore unlikely to predict the true value at higher redshift.

An additional set of simulations was run to mimic lines of sight emanating from galaxies like those in the KBSS. In these simulations, H I absorbers were added within 700 km s^{-1} of $z_{\rm em}$ according to the CGM broken power law fit to $f(N,X)$. The remainder of the line of sight is drawn from the IGM distribution. These simulations produce the expected attenuation due to the CGM for sources within galaxies similar to those in our spectroscopic sample ($0.25 < L/L^* < 3$, $\langle z \rangle = 2.3$).

In addition to the simulated forest spectra that include both line and continuum opacity, we created another set where the opacity from individual Lyman lines is not included, leaving only Lyman continuum absorption. To measure the $\lambda_{\rm mfp}$, the principal concern is the fraction of photons absorbed while they are at ionizing energies. Once photons redshift out of the hydrogen-ionizing band, for the purpose of calculating the mean free path, their subsequent absorption or transmission is immaterial. For this reason, spectra with line and continuum opacity are useful for correcting observed objects for attenuation along the line of sight through the IGM (as will be discussed in Section 6.4), while the "continuum absorption only" spectra can be used to measure $\lambda_{\rm mfp}$ directly (as discussed in Section 6.2).

Figure 11 shows the average of 10,000 realizations of each Monte Carlo run plotted in the rest-frame of the emitter. The solid curves show the runs with Lyman line opacity included, and the dotted and dash-dotted lines show the MC runs that only consider the continuum opacity. The various colors pertain to either IGM ONLY (blue and purple) or IGM+CGM (red and orange) runs (note the presence of Lyα, β, γ, etc. absorption due to the CGM of the emitter in the red spectrum). Note that at the Lyman limit, the spectra that include CGM absorption show a sharper drop in transmission. This is due to LyC absorption from high-$N_{\rm HI}$ absorbers found in the circumgalactic gas.

6.2. *Measuring* $\lambda_{\rm mfp}$

To measure $\lambda_{\rm mfp}$ from the Monte Carlo calculation, we use the spectra that include continuum absorption only.[15] We search the average spectrum for the location where the LyC only spectra reach a transmission equivalent to e^{-1}, as marked by the vertical dash-dotted lines in Figure 12. The rest wavelengths from Figure 12 can be converted into the physical distance traveled by a photon between $z_{\rm em} = 2.4$, the emission redshift of the spectra, and $z_{\rm mfp}$ where:

$$z_{\rm mfp} = \frac{\lambda_e - \lambda_{\rm LL}}{\lambda_{\rm LL}}(1+z_{\rm em}) + z_{\rm em} \tag{25}$$

where $\lambda_{\rm LL} = 911.75$ Å is the wavelength of the Lyman limit and λ_e is the rest wavelength of the spectrum at which the transmission equals e^{-1}. To calculate $\lambda_{\rm mfp}$, we then compute the physical distance traveled by the photon through the expanding universe, given by the integral over the proper line element:

$$\lambda_{\rm mfp} \equiv \int_{z_{\rm mfp}}^{z_{\rm em}} \frac{1}{1+z}\frac{c}{H(z)} dz \tag{26}$$

where c is the speed of light and

$$H(z) = H_0 \sqrt{\Omega_\Lambda + \Omega_{\rm m}(1+z)^3}. \tag{27}$$

Using the averaged LyC only spectra shown in Figure 12, we calculate the $\lambda_{\rm mfp}$ through the IGM assuming $z_{\rm em} = 2.4$ to be 147 pMpc (or $\Delta z = 0.359$). If instead the CGM is included (as would be the case for LyC photons emanating from a galaxy), the estimated $\lambda_{\rm mfp}$ falls to 120 pMpc

[14] Considered in the redshift range $2.0 < z < 2.8$, varying γ by 1σ results in less than a 6% change in the incidence of LLSs and therefore less than a 3% change in $\lambda_{\rm mfp}$.

[15] Line opacity only reduces the transmission of non-ionizing photons which are unimportant for consideration of $\lambda_{\rm mfp}$.

(Δz = 0.301), $\sim$20% shorter. Note that the values of $\lambda_{\rm mfp}$ found in this sections using the MC simulations are similar to those estimated using the analytic approximations in Section 5 ($\lambda^{\rm static}_{\rm mfp,\ IGM}$= 122.8 pMpc, $\lambda^{\rm static}_{\rm mfp}$= 109.3 pMpc). This provides confidence that the conclusions of Section 5 hold even when a more rigorous treatment is considered.

6.3. *Computing the mean free path uncertainties*

To calculate the uncertainties in the $\lambda_{\rm mfp}$ measurement made in Sections 6.2, we must estimate the error in the opacity that originates from the Lyα forest absorbers ($\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$) and from LLSs ($\log(N_{\rm HI}/{\rm cm}^{-2}) > 17.2$). The errors associated with each must be estimated separately because the constraints on the incidence of these absorbers come from different surveys.

To estimate the uncertainty associated with LLSs, we return to the measured value of $dn_{\rm LLS}/dX = 0.52 \pm 0.08$ (Section 5). Using the method outlined in Section 5, we employ a fit to the frequency distribution to calculate the opacity, κ. Once again, we assume that an extrapolation of the broken power law fit to the slope of the high-$N_{\rm HI}$ absorbers holds in the LLS regime. We then calculate the normalization of the frequency distribution that reproduces $dn_{\rm LLS}/dX$ and its upper and lower limits. These normalizations and the assumed slope can then be used to measure $\kappa_{\rm LLS}$ and an approximate error. We find $\kappa_{\rm LLS} = 0.0043 \pm 0.0006$ Mpc^{-1}, a 15% uncertainty. If we allow the value of β to vary within the $1-\sigma$ confidence region (see Table 6), the uncertainty in our estimate of $\kappa_{\rm LLS}$ would be larger, $\sim 20-25$%.

To estimate the $\lambda_{\rm mfp}$ uncertainties from absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ we employ the measurement of $\kappa_{\rm data}$ from Section 5. We thus assume that the uncertainty is dominated by sample variance rather than the uncertainties in the model. However, before using this value, we verify that the opacity (and its uncertainty) from $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ absorbers measured from the data are reproduced in the Monte Carlo simulations using the assumed power-law fits.

To estimate the opacity in the MC runs from forest absorbers, we ran another set of MC simulations of LyC only opacity. In this case, we included only absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$, excluding the contribution from LLSs. Using the ensemble of model spectra, we performed a jack-knife test with a sample size of 15 sightlines, designed to match the number of lines of sight in the KBSS QSO sample. Each set of 15 randomly-drawn spectra was averaged and the maximum attenuation with respect to the continuum was measured and recorded. The rest-frame wavelength at the position of maximum attenuation was converted into a proper distance at the emission redshift using equation 26. The average value of κ along those 15 lines of sight was taken to be the natural logarithm of the minimum flux divided by physical distance. The redshift range used in the simulations is $2.0 < z < 2.8$ in order to reproduce the range of the KBSS QSO sightline sample.

The jack-knife technique was applied 10,000 times to produce a distribution of the forest opacity, $\kappa_{\rm for}$, as shown in Figure 13. An estimate for the error in $\kappa_{\rm for}$ was then derived from the moments of the distribution. The resulting value of $\kappa_{\rm for} = 0.0037 \pm 0.0005$ is in good agreement with $\kappa_{\rm data} = 0.0039 \pm 0.0005$ calculated in Section 5, providing confidence that the MC simulations well reproduce both the mean value and the dispersion in $\kappa_{\rm for}$ for a sample of the same size as the KBSS.

Table 8
Comparison of the $\lambda_{\rm mfp}$ derived from the LyC and the renormalized forest spectra

$z_{\rm em}$	$\lambda_{\rm mfp}$ LyC [pMpc] IGM ONLY	$\lambda_{\rm mfp}\langle 1-D_B\rangle$ [pMpc] IGM ONLY	$\lambda_{\rm mfp}$ LyC [pMpc] IGM+CGM	$\lambda_{\rm mfp}\langle 1-D_B\rangle$ [pMpc] IGM+CGM
2.2	191.8	185.9	160.2	151.4
2.4	146.5	135.7	120.0	113.4
2.6	115.7	106.8	93.7	85.6
2.8	91.9	85.1	73.5	66.0

With an estimate of $\kappa_{\rm LLS}$, $\sigma_{\kappa,\rm LLS}$ as well as $\kappa_{\rm for}$ and $\sigma_{\kappa,\rm for}$, it is straightforward to calculate the error in the $\lambda_{\rm mfp}$. If

$$\lambda_{\rm mfp} = \frac{1}{\kappa_{\rm LLS} + \kappa_{\rm for}} \tag{28}$$

(see equation 23) then the uncertainty on $\lambda_{\rm mfp}$ is

$$\sigma_{\rm mfp} = \frac{\sqrt{\sigma^2_{\kappa,\rm LLS} + \sigma^2_{\kappa,\rm for}}}{(\kappa_{\rm LLS} + \kappa_{\rm for})^2} \tag{29}$$

where $\sigma_{\kappa,\rm LLS}$ and $\sigma_{\kappa,\rm for}$ are the uncertainties of $\kappa_{\rm LLS}$ and $\kappa_{\rm for}$ respectively. Employing these equations, we find a typical uncertainty in $\lambda_{\rm mfp}$ of ~ 15 Mpc, which we adopt for both the IGM and IMG+CGM $\lambda_{\rm mfp}$ measurements. Since the contribution of LLSs to the opacity in the CGM is poorly constrained, the error on the CGM measurements is likely somewhat larger; however since the CGM contributes < 20% of the opacity, the effect on the net uncertainty is modest. We adopt this uncertainty for the measurement of $\lambda_{\rm mfp}$ at the mean redshift of the sample.

6.4. *Simulated sightlines with line and continuum opacity*

For some applications, it is more useful to know the expected attenuation along an ensemble of sight lines including line opacity. For instance, in studies of the escape fraction of LyC photons (Shapley et al. 2006; Iwata et al. 2009; Bridge et al. 2010; Siana et al. 2010; Nestor et al. 2011; Steidel et al. in prep) a correction must be made for the fraction of the attenuation at a given observed wavelength due to (non-ionizing) line blanketing. For these applications, our models including CGM attenuation are ideal. The sample of galaxies used in our CGM study have impact parameters of $50-300$ pkpc from the QSO line of sight, and so our MC simulations which include opacity from the CGM effectively model the typical gas distribution at radii larger than 50 kpc - and hence the attenuation suffered by photons after they leave the *ISM* of the galaxy, but before they leave the large gas densities associated with the CGM.

Using the sets of simulated spectra with $z_{\rm em}$ = 2.4 that include Lyman line opacity for the IGM ONLY and the IGM+CGM, we measure the distribution of transmission of emitted 880-910 Å photons through the IGM and the IGM + CGM; see Figure 14. The transmission is generally lower along the IGM+CGM sight lines. The peak in the transmission PDF near ~ 0.75 corresponds to the average transmission within the Lyβ forest which is < 1 due to line blanketing from both higher Lyman series lines as well as Lyα absorption from lower redshift.

In Figure 12, the solid curves show the average spectrum of 10,000 MC runs that include line opacity. We use $\langle 1-D_B\rangle$, the average transmission in the region between 920 -1015 Å in the

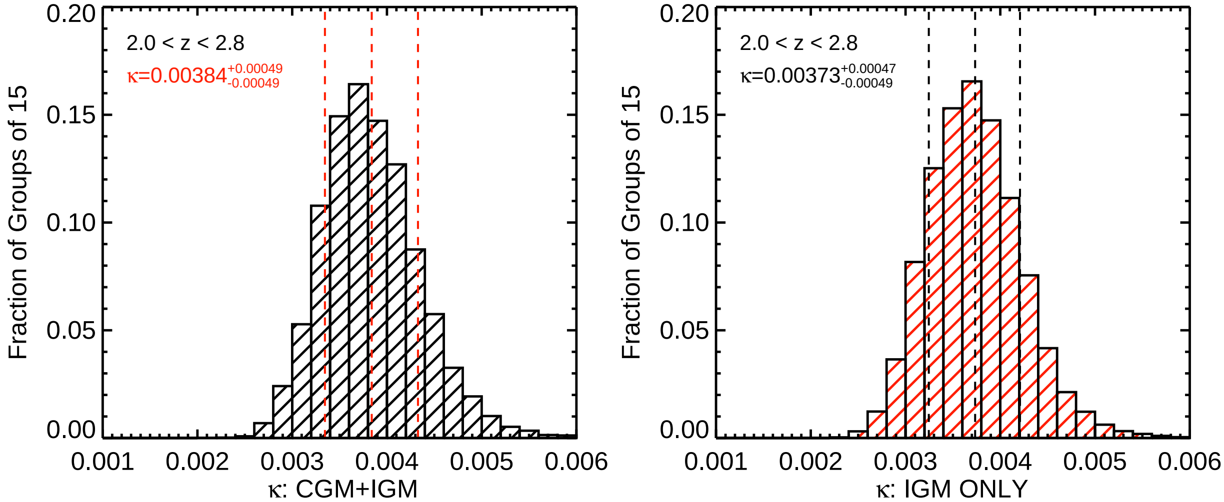

Figure 13. The variation in the measured opacity in groups of 15 sightlines due to all absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$. The simulations used in this plot are those with emission redshift $z = 2.8$. These show the results of 10,000 jack-knife tests in which 15 spectra were randomly drawn from the "LyC ONLY" Monte Carlo runs in which only absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ were included. See Section 6.3. The dotted vertical lines mark the $1-\sigma$ uncertainties in κ.

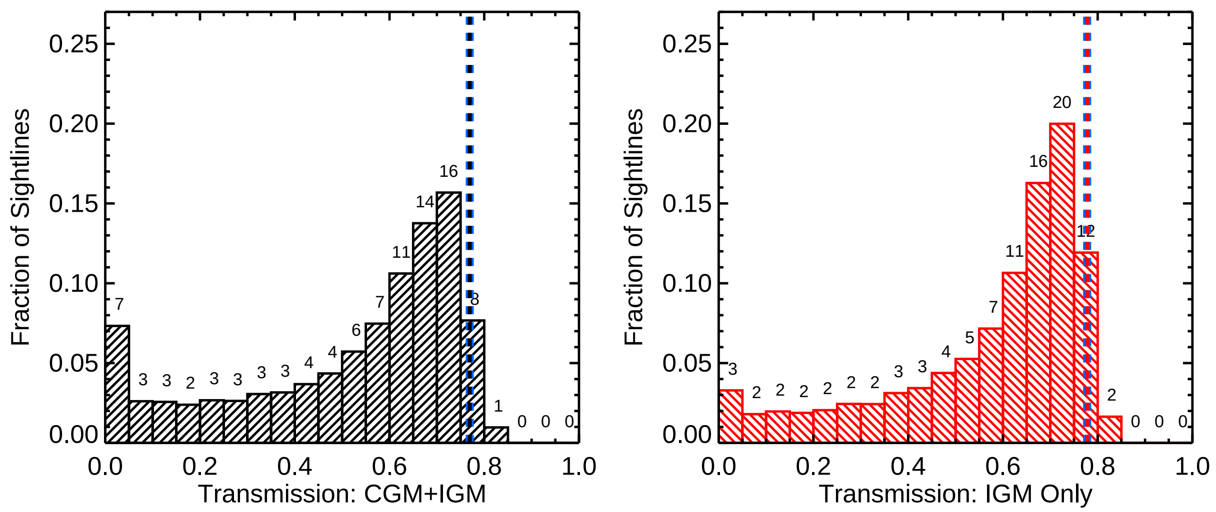

Figure 14. The distribution of transmissions from each sightline averaged over the band pass from 880-910 Å for the the IGM+CGM Monte Carlo simulations (left panel) and IGM only Monte Carlo simulations (right panel) that include line opacity. The simulations used for these figures are those with an emission redshift of $z = 2.4$. The (blue) dashed vertical line marks the average transmission in the Ly$\beta,\gamma,+$ forest $\langle 1-D_B \rangle$.

rest frame of the source (the region shown in the yellow box in Figure 11; Oke & Korycansky 1982) to "re-normalize" the continuum against which the LyC opacity will be measured. Using the renormalized spectrum as an approximation to the "LyC only" MC spectra and measuring $\lambda_{\rm mfp}$ as described in Section 6.2, leads to $\lambda_{\rm mfp}$ values $5-10\%$ smaller than the "true" values. Table 8 compares the results of the MC simulations including line opacity (columns labeled $\langle 1-D_B \rangle$) with those based on the continuum only spectra (columns labeled LyC).

6.5. *Comparison with previous $\lambda_{\rm mfp}$ measurements*

In the previous sections, we derived the value of the mean free path $\lambda_{\rm mfp}$ to LyC photons traveling through the IGM to be:

$$\text{IGM ONLY: } \lambda_{\rm mfp}(z_{\rm em} = 2.4) = 147 \pm 15\,{\rm Mpc}. \qquad (30)$$

If instead, we assume such LyC photons emanate from galaxies similar to one of those in our spectroscopic sample, we find a value for $\lambda_{\rm mfp}$ including the CGM opacity to be:

$$\text{IGM+CGM: } \lambda_{\rm mfp}(z_{\rm em} = 2.4) = 121 \pm 15\,{\rm Mpc}. \qquad (31)$$

One important advance of the estimate of $\lambda_{\rm mfp}$ made in this paper is that all parts of the column density distribution have now been measured at the same mean redshift. This has not been possible previously due to the difficulty of collecting appropriate statistical samples for the various column density regions of the frequency distribution.

In this section, we compare these values to previous estimates from the literature. One caveat is that the measurements made in this paper typically differ in mean redshift from the samples on which previous $\lambda_{\rm mfp}$ measurements have relied. As such, an extrapolation of the literature measurements or our own measurements are generally necessary in order to

compare the results, and therefore these comparisons are more dependent on the value of γ assumed. Nevertheless, in this section we compare to previous measurements for completeness. Figure 15 gives a graphical summary of the results of this comparison.

In general, the differences between our IGM ONLY $\lambda_{\rm mfp}$ results and those of other authors (especially those measured at similar z) underscore the degree to which $\lambda_{\rm mfp}$ is sensitive to changes in $f(N,X)$, highlighting the importance of the precise measurement of the intermediate $N_{\rm HI}$ systems made here.

One other caveat is the small degree of incompleteness which may be present in the $z \lesssim 2.4$, $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 15.5$ portion of our absorber sample. As described in Appendix B, this bias will only act to *shorten* the inferred value of $\lambda_{\rm mfp}$, pushing it to values farther below those inferred by other authors.

The mean free path and its evolution with redshift are generally parameterized as:

$$\lambda_{\rm mfp} = \lambda_{\rm mfp,0}\,(1+z)^{\xi} \tag{32}$$

where ξ captures the cosmology (as parameterized in equation 26) as well as the evolution in the number of absorbers as a function of redshift, dn/dz (described in Section 6.1). For the Monte Carlo simulations described in this section, the redshift evolution of $\lambda_{\rm mfp}$ is fixed by the assumed value of γ. We emphasize that the redshift evolution is only constrained by data with $2.0 \lesssim z \lesssim 3.0$, and as such the extrapolation of our results above to $z >> 3$ is ill-advised.

The redshift evolution of $\lambda_{\rm mfp}$, parameterized as a function of the emission redshift of the QSO, $z_{\rm em}$, returned by the MC simulations is represented by the solid blue curves in Figure 15 and is given by:

$$\text{IGM ONLY: } \lambda_{\rm mfp}(z_{\rm em}) = 147\left(\frac{1+z_{\rm em}}{3.4}\right)^{-4.3}\ \text{pMpc} \tag{33}$$

and

$$\text{IGM+CGM: } \lambda_{\rm mfp}(z_{\rm em}) = 121\left(\frac{1+z_{\rm em}}{3.4}\right)^{-4.5}\ \text{pMpc.} \tag{34}$$

The first estimate of $\lambda_{\rm mfp}$ was made by Madau et al. (1999) who found:

$$\lambda_{\rm mfp} = 42\left(\frac{1+z}{4.0}\right)^{-4.5}\ \text{Mpc} = 87\left(\frac{1+z}{3.4}\right)^{-4.5}\ \text{Mpc} \tag{35}$$

using a simple approximation for $f(N,X)$ with $\beta = 1.5$ and $\gamma = 2$ where their Einstein-de Sitter cosmology has been converted to the cosmology used in this paper.

Faucher-Giguère et al. (2008) published an updated estimate to the $\lambda_{\rm mfp}$ assuming $\beta = 1.39$ as measured by Misawa et al. (2007). Misawa et al. (2007) measured the frequency distribution close to high-$N_{\rm HI}$ absorbers, and thus their sample is similar in slope to our CGM $f(N,X)$ distribution. However, the normalization chosen by Faucher-Giguère et al. (2008) was computed using LLS statistics for the full IGM, thereby closely matching our LLS points but somewhat underpredicting the incidence of lower-$N_{\rm HI}$ absorbers resulting in slightly larger values of $\lambda_{\rm mfp}$ at a given z:

$$\lambda_{\rm mfp} = 85\left(\frac{1+z}{4}\right)^{-4}\ \text{Mpc} = 160\left(\frac{1+z}{3.4}\right)^{-4}\ \text{Mpc} \tag{36}$$

Prochaska et al. (2009) estimated $\lambda_{\rm mfp}$ using stacked QSO

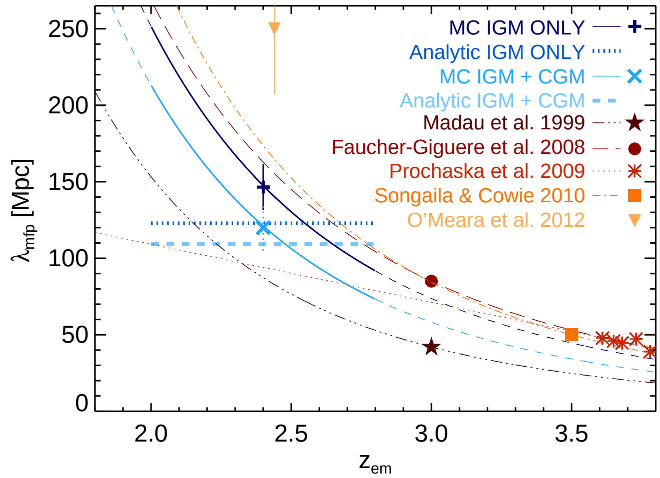

Figure 15. The $\lambda_{\rm mfp}$ as a function of $z_{\rm em}$ including previous estimates of the $\lambda_{\rm mfp}$ from the literature. The blue points and curves refer to values measured in this paper whereas estimates from the literature are plotted in shades of red. The blue cross and X and their corresponding curves show the values of $\lambda_{\rm mfp}$ calculated from the MC spectra for the IGM ONLY and IGM+CGM runs respectively. The dotted and dashed blue horizontal line show the results of the analytic calculation in Section 5 excluding and including the CGM respectively. The dark red filled star and the dash-dotted curve that intersects it are the suggested values of $\lambda_{\rm mfp}$ from Madau et al. (1999). The red filled circle and dashed curve are the results of Faucher-Giguère et al. (2008). Prochaska et al. (2009) suggested the values shown with the red/orange asterisks and the dotted line. The measurements of Songaila & Cowie (2010) are shown in the orange filled square and the dash-dotted curve that intersects it. The results of O'Meara et al. (2012) are shown in the orange filled triangle.

spectra to measure the LyC opacity. Notably, such a method would include any contribution from the CGM of their QSOs. They found

$$\lambda_{\rm mfp} = 48.4 - 38.0(z-3.6)\ \text{Mpc} = 94 - 38.0(z-2.4)\ \text{Mpc} \tag{37}$$

somewhat lower than our measurement at $z = 2.4$ and with very different evolution with redshift than indicated by our data.

Songaila & Cowie (2010) revisited the mean free path calculation after compiling a large sample of $z < 1$ and $z > 4$ LLS. They found

$$\lambda_{\rm mfp} = 50\left(\frac{1+z}{4.5}\right)^{-4.44^{+0.36}_{-0.32}}\ \text{Mpc} = 170\left(\frac{1+z}{3.4}\right)^{-4.44^{+0.36}_{-0.32}}\ \text{Mpc} \tag{38}$$

using $\beta = 1.3$. They also offer a variety of normalizations for different β values, including 130 Mpc for $\beta = 1.5$ at $z = 2.4$, in general agreement with our measured "IGM ONLY" value of $\lambda_{\rm mfp}$.

Recently O'Meara et al. (2012) presented a calculation of $\lambda_{\rm mfp}$ using their new measurements of the incidence of LLSs at $z \sim 2$. Their estimates of the incidence of absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ leads to a lower value for the opacity as described in Section 5.1. They found

$$\lambda_{\rm mfp} = 250 \pm 43\ \text{Mpc} \tag{39}$$

at $z = 2.44$, nearly a factor of 2 larger than our estimate and significantly larger than the extrapolations of others' measurements at higher redshifts.

We note that for a steeper evolution in redshift as found

in Songaila & Cowie (2010) and Prochaska et al. (2010), the general agreement between the extrapolation of the presented measurements at $\langle z \rangle = 2.4$ to $z > 3$ would result in a more discrepant value significantly below the measurements at $z > 3.5$ by Songaila & Cowie (2010) and Prochaska et al. (2009).

To summarize, the calculation described in Section 6 offers three significant improvements over previous measurements. First, all portions of $f(N,X)$ have been measured at the same mean redshift, avoiding the need to extrapolate any part of $f(N,X)$ to a different redshift using uncertain values of γ. A second advance is the accuracy of our Voigt profile fitting technique, achieved by including higher-order Lyman series transitions, yielding the best measurement of the power law slope, β, for absorbers with saturated Lyα profiles, $[14 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2]$. Because such absorbers account for ∼half of the opacity to LyC photons in the IGM, underestimating their contribution would result in significantly underestimating the total opacity of the IGM.

The third major improvement with respect to previous studies is the inclusion of absorption from the CGM which again decreases the measured value of $\lambda_{\rm mfp}$. If LyC photons are presumed to originate within galaxies similar to those in the KBSS sample, then every hydrogen-ionizing photon must first propagate through the CGM. As shown by Rudie et al. (2012), regions of the IGM near galaxies have a much higher incidence of high-$N_{\rm HI}$ absorbers than the average IGM with the effect (again) of reducing $\lambda_{\rm mfp}$ relative to previous measurements.

6.6. *Implications of the higher opacity of the IGM+CGM*

A shorter mean free path has implications for the emissivity and demographics of ionizing sources at high redshift, as well as for calculations of the metagalactic UV radiation field. Briefly, if one considers the specific intensity of the UV background (J_ν, see e.g. Scott et al. 2000) or the photoionization rate (Γ, see e.g. Faucher-Giguère et al. 2008) to be well determined, a shorter mean free path directly implies a requirement for correspondingly larger ionizing emissivity of sources (ϵ_ν). Following Faucher-Giguère et al. (2008),

$$J_\nu(z) \approx \frac{1}{4\pi} \lambda_{\rm mfp}(\nu,z)\ \epsilon_\nu(z) \qquad (40)$$

and

$$\Gamma \propto J_\nu. \qquad (41)$$

Conversely, for a measured value of ϵ_ν from a given population of sources, a shorter $\lambda_{\rm mfp}$ implies a smaller contribution to both the specific intensity and the photoionization rate (Γ). Higher intergalactic opacity also requires that the background be produced by sources within a smaller volume, possibly leading to greater spatial variation in the radiation field intensity.

LyC escape fraction studies have often found very few detections of ionizing photons escaping from galaxies (Shapley et al. 2006; Bridge et al. 2010; Siana et al. 2010; however see Iwata et al. 2009, Nestor et al. 2011, and Steidel et al. in prep). For LyC studies of galaxies at high-z, the inclusion of the CGM is an important effect. Considering Figure 14, note that the number of galaxies with ∼zero transmission increases by a factor of 2 when the CGM is included. If the redshift evolution of the CGM is similar to that measured for the IGM, the effect of the CGM will become more pronounced at higher-z where the incidence of high-$N_{\rm HI}$ systems in the CGM will be much higher.

Higher IGM and CGM opacities mean that there may be more sources contributing to the metagalactic UV radiation field than are implied by the number of actual LyC detections; on the other hand, the assumption of a $\lambda_{\rm mfp}$ that is too large may cause one to over-estimate the contribution to J_ν or Γ made by a population of detected LyC sources. The latter effect may reduce the tension between the apparent over-production of ionizing photons by observed LBGs at $z \sim 3$ (Nestor et al. 2011) relative to estimates of the total photoionization rate produced by all sources (Bolton et al. 2005; Faucher-Giguère et al. 2008).

7. SUMMARY

Using a sample of 15 high-S/N, high-resolution QSO spectra drawn from the KBSS, we have produced the largest catalog of H I absorbers fit with Voigt profiles to Lyα and at least one higher order Lyman series transition. This analysis enables the first statistically rigorous measurements of the frequency distribution of H I absorbers with $14 \lesssim \log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 17$. In Section 3, we showed that the frequency of absorbers as a function of $N_{\rm HI}$ is well-parameterized by a single power-law from $13.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ with a maximum likelihood index of $\beta = 1.65 \pm 0.02$ and a normalization per unit pathlength of $\log(C_{\rm HI}) = 10.32$.

In Section 4, we measured the frequency distribution within 300 pkpc and both 300 km s^{-1} and 700 km s^{-1} of galaxies in the KBSS sample. We showed that the frequency of absorbers near galaxies is significantly higher and that the power-law index is shallower compared with that of the IGM, meaning there are disproportionately more high-$N_{\rm HI}$ systems than low. These findings are discussed further by Rudie et al. (2012).

Section 5 examined statistically the total opacity of the IGM and CGM to hydrogen-ionizing Lyman continuum (LyC) photons. This section presented measurements of the opacity due to absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ with direct constraints from the data. We also found that a single MLE power-law fit over the full range of $N_{\rm HI}$ does not well reproduce this opacity, and also fails to reproduce the incidence of LLSs. We therefore adopted a broken power law fit that reproduces both the opacity measured within the data from absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ as well as the observed incidence of LLSs. The parameters of these fits can be found in Table 5.

We also measured the fractional LyC opacity in bins of $N_{\rm HI}$. We found that 48% of the opacity within one mean free path, $\lambda_{\rm mfp}$, is contributed by absorbers with $N_{\rm HI} < 10^{17.2}$ cm^{-2}, the vast majority of which results from absorbers with $14 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ whose frequency was poorly measured prior to this work.

In Section 6, we used Monte Carlo simulations to measure the value of the mean free path ($\lambda_{\rm mfp}$) of LyC photons at $\langle z \rangle = 2.4$. We measured both the $\lambda_{\rm mfp}$ of photons through the IGM (as has been considered many previous times) as well as the $\lambda_{\rm mfp}$ of photons formed in galaxies similar to those in the KBSS - photons that must first transverse the CGM of these galaxies before reaching the lower-opacity IGM. We found $\lambda_{\rm mfp}(z = 2.4) = 147 \pm 15$ Mpc for simulations including only IGM opacity and $\lambda_{\rm mfp}(z = 2.4) = 121 \pm 15$ Mpc when we include the CGM. These values of $\lambda_{\rm mfp}$ are lower than most previous estimates, which have not included the CGM and have generally underestimated the contribution of intermediate-$N_{\rm HI}$ absorbers. We also quantified the effect of the CGM on the distribution of transmission through random sight lines, rele-

vant to measurements of the escape fractions of LyC photons from such galactic sources at high-z.

We note that the $\lambda_{\rm mfp}$ value including attenuation from the CGM is relevant only if the dominant sources of LyC photons are galaxies similar to the UV color-selected galaxies in our sample ($0.25 < L/L^{*} < 3.0$ at $\langle z \rangle = 2.3$) - for other galaxies, these measurements are an approximation. However, in all cases the general statement holds that $\lambda_{\rm mfp}$ and related statistics are affected by the $\sim$Mpc-scale gaseous environments of the sources that contribute significantly to the metagalactic background.

The IGM and CGM opacities measured in this paper have significant implications for studies of ionizing sources at high redshift and for estimates of the metagalactic UV background at $z \approx 2-3$.

The authors wish to thank Claude-André Faucher-Giguère for his careful reading of the draft and pertinent comments. We also thank Olivera Rakic for her contributions to the reduction of the QSO data set and for her helpful advice. The authors wish to acknowledge Ryan Cooke who contributed the fits to the damped profiles in our QSO spectra. Our thanks to Bob Carswell for his assistance with VPFIT.

We wish to acknowledge the staff of the the W.M. Keck Observatory whose efforts insure the telescopes and instruments perform reliably. Further, we extend our gratitude to those of Hawaiian ancestry on whose sacred mountain we are privileged to be guests.

This work has been supported by the US National Science Foundation through grants AST-0606912 and AST- 0908805. CCS acknowledges additional support from the John D. and Catherine T. MacArthur Foundation and the Peter and Patricia Gruber Foundation. This research has made use of the Keck Observatory Archive (KOA), which is operated by the W. M. Keck Observatory and the NASA Exoplanet Science Institute (NExScI), under contract with the National Aeronautics and Space Administration.

REFERENCES

Adelberger, K. L., Steidel, C. C., Pettini, M., Shapley, A. E., Reddy, N. A., & Erb, D. K. 2005, ApJ, 619, 697
Altay, G., Theuns, T., Schaye, J., Crighton, N. H. M., & Dalla Vecchia, C. 2011, ApJ, 737, L37
Bahcall, J. N. & Peebles, P. J. E. 1969, ApJ, 156, L7
Becker, G. D., Rauch, M., & Sargent, W. L. W. 2009, ApJ, 698, 1010
Becker, G. D., Sargent, W. L. W., Rauch, M., & Calverley, A. P. 2011, ApJ, 735, 93
Becker, G. D., Sargent, W. L. W., Rauch, M., & Carswell, R. F. 2012, ApJ, 744, 91
Bolton, J. S., Haehnelt, M. G., Viel, M., & Springel, V. 2005, MNRAS, 357, 1178
Bridge, C. R., Teplitz, H. I., Siana, B., Scarlata, C., Conselice, C. J., Ferguson, H. C., Brown, T. M., Salvato, M., Rudie, G. C., de Mello, D. F., Colbert, J., Gardner, J. P., Giavalisco, M., & Armus, L. 2010, ApJ, 720, 465
Carswell, R. F., Morton, D. C., Smith, M. G., Stockton, A. N., Turnshek, D. A., & Weymann, R. J. 1984, ApJ, 278, 486
Cen, R., Miralda-Escudé, J., Ostriker, J. P., & Rauch, M. 1994, ApJ, 437, L9
Clauset, A., Rohilla Shalizi, C., & Newman, M. E. J. 2007, arXiv: 0706.1062
Conroy, C., Shapley, A. E., Tinker, J. L., Santos, M. R., & Lemson, G. 2008, ApJ, 679, 1192
Corbelli, E., Salpeter, E. E., & Bandiera, R. 2001, ApJ, 550, 26
Davé, R., Hellsten, U., Hernquist, L., Katz, N., & Weinberg, D. H. 1998, ApJ, 509, 661
Davé, R. & Tripp, T. M. 2001, ApJ, 553, 528
Ellison, S. L., Lewis, G. F., Pettini, M., Sargent, W. L. W., Chaffee, F. H., Foltz, C. B., Rauch, M., & Irwin, M. J. 1999, Publications of the Astronomical Society of the Pacific, 111, pp. 946
Erb, D. K., Steidel, C. C., Shapley, A. E., Pettini, M., Reddy, N. A., & Adelberger, K. L. 2006, ApJ, 647, 128
Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, ApJ, 688, 85
Gunn, J. E. & Peterson, B. A. 1965, ApJ, 142, 1633
Hernquist, L., Katz, N., Weinberg, D. H., & Miralda-Escudé, J. 1996, ApJ, 457, L51+
Hu, E. M., Kim, T.-S., Cowie, L. L., Songaila, A., & Rauch, M. 1995, AJ, 110, 1526
Iwata, I., Inoue, A. K., Matsuda, Y., Furusawa, H., Hayashino, T., Kousai, K., Akiyama, M., Yamada, T., Burgarella, D., & Deharveng, J.-M. 2009, ApJ, 692, 1287
Janknecht, E., Reimers, D., Lopez, S., & Tytler, D. 2006, A&A, 458, 427
Katz, N., Weinberg, D. H., Hernquist, L., & Miralda-Escude, J. 1996, ApJ, 457, L57
Kim, T.-S., Carswell, R. F., Cristiani, S., D'Odorico, S., & Giallongo, E. 2002, MNRAS, 335, 555
Lehner, N., Savage, B. D., Richter, P., Sembach, K. R., Tripp, T. M., & Wakker, B. P. 2007, ApJ, 658, 680
Madau, P., Haardt, F., & Rees, M. J. 1999, ApJ, 514, 648
McQuinn, M., Oh, S. P., & Faucher-Giguère, C.-A. 2011, ApJ, 743, 82
Meiksin, A. A. 2009, Reviews of Modern Physics, 81, 1405
Misawa, T., Tytler, D., Iye, M., Kirkman, D., Suzuki, N., Lubin, D., & Kashikawa, N. 2007, AJ, 134, 1634
Murakami, I. & Ikeuchi, S. 1990, PASJ, 42, L11
Nestor, D. B., Shapley, A. E., Steidel, C. C., & Siana, B. 2011, ApJ, 736, 18
Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009, A&A, 505, 1087
Oke, J. B. & Korycansky, D. G. 1982, ApJ, 255, 11
O'Meara, J. M., Prochaska, J. X., Burles, S., Prochter, G., Bernstein, R. A., & Burgess, K. M. 2007, ApJ, 656, 666
O'Meara, J. M., Prochaska, J. X., Worseck, G., Chen, H.-W., & Madau, P. 2012, arXiv: 1204.3093
Paresce, F., McKee, C. F., & Bowyer, S. 1980, ApJ, 240, 387
Penton, S. V., Stocke, J. T., & Shull, J. M. 2004, ApJS, 152, 29
Petitjean, P., Bergeron, J., & Puget, J. L. 1992, A&A, 265, 375
Petitjean, P., Webb, J. K., Rauch, M., Carswell, R. F., & Lanzetta, K. 1993, MNRAS, 262, 499
Prochaska, J. X., O'Meara, J. M., & Worseck, G. 2010, ApJ, 718, 392
Prochaska, J. X., Worseck, G., & O'Meara, J. M. 2009, ApJ, 705, L113
Rauch, M. 1998, ARA&A, 36, 267
Reddy, N., Dickinson, M., Elbaz, D., Morrison, G., Giavalisco, M., Ivison, R., Papovich, C., Scott, D., Buat, V., Burgarella, D., Charmandaris, V., Daddi, E., Magdis, G., Murphy, E., Altieri, B., Aussel, H., Dannerbauer, H., Dasyra, K., Hwang, H. S., Kartaltepe, J., Leiton, R., Magnelli, B., & Popesso, P. 2012, ApJ, 744, 154
Reddy, N. A. & Steidel, C. C. 2009, ApJ, 692, 778
Reddy, N. A., Steidel, C. C., Pettini, M., Adelberger, K. L., Shapley, A. E., Erb, D. K., & Dickinson, M. 2008, ApJS, 175, 48
Rees, M. J. 1986, MNRAS, 218, 25P
Ribaudo, J., Lehner, N., & Howk, J. C. 2011, ApJ, 736, 42
Rudie, G. C., Steidel, C. C., Trainor, R. F., Rakic, O., Bogosavljević, M., Pettini, M., Reddy, N., Shapley, A. E., Erb, D. K., & Law, D. R. 2012, ApJ, 750, 67
Ryan-Weber, E. V., Pettini, M., Madau, P., & Zych, B. J. 2009, MNRAS, 395, 1476
Sargent, W. L. W., Steidel, C. C., & Boksenberg, A. 1989, ApJS, 69, 703
Schaye, J. 2001, ApJ, 559, 507
Schaye, J., Aguirre, A., Kim, T.-S., Theuns, T., Rauch, M., & Sargent, W. L. W. 2003, ApJ, 596, 768
Scott, J., Bechtold, J., Dobrzycki, A., & Kulkarni, V. P. 2000, ApJS, 130, 67
Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., & Erb, D. K. 2006, ApJ, 651, 688
Siana, B., Teplitz, H. I., Ferguson, H. C., Brown, T. M., Giavalisco, M., Dickinson, M., Chary, R.-R., de Mello, D. F., Conselice, C. J., Bridge, C. R., Gardner, J. P., Colbert, J. W., & Scarlata, C. 2010, ApJ, 723, 241
Simcoe, R. A. 2011, ApJ, 738, 159
Simcoe, R. A., Cooksey, K. L., Matejek, M., Burgasser, A. J., Bochanski, J., Lovegrove, E., Bernstein, R. A., Pipher, J. L., Forrest, W. J., McMurtry, C., Fan, X., & O'Meara, J. 2011, ApJ, 743, 21
Simcoe, R. A., Sargent, W. L. W., & Rauch, M. 2004, ApJ, 606, 92
Songaila, A. & Cowie, L. L. 1996, AJ, 112, 335
—. 2010, ApJ, 721, 1448

Stengler-Larrea, E. A., Boksenberg, A., Steidel, C. C., Sargent, W. L. W., Bahcall, J. N., Bergeron, J., Hartig, G. F., Jannuzi, B. T., Kirhakos, S., Savage, B. D., Schneider, D. P., Turnshek, D. A., & Weymann, R. J. 1995, ApJ, 444, 64
Trainor, R. F. & Steidel, C. C. 2012, ApJ, 752, 39
Tytler, D. 1982, Nature, 298, 427
—. 1987, ApJ, 321, 49
Umemura, M. & Ikeuchi, S. 1985, ApJ, 299, 583
Vogt, S. S., Allen, S. L., Bigelow, B. C., Bresee, L., Brown, B., Cantrall, T., Conrad, A., Couture, M., Delaney, C., Epps, H. W., Hilyard, D., Hilyard, D. F., Horn, E., Jern, N., Kanto, D., Keane, M. J., Kibrick, R. I., Lewis, J. W., Osborne, J., Pardeilhan, G. H., Pfister, T., Ricketts, T., Robinson, L. B., Stover, R. J., Tucker, D., Ward, J., & Wei, M. Z. 1994, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 2198, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. D. L. Crawford & E. R. Craine, 362–+
Zheng, Z. & Miralda-Escudé, J. 2002, ApJ, 568, L71

APPENDIX

A. FITS TO HIGH-$N_{\rm HI}$ ABSORBERS

As discussed in §1, for $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 14.5$, the Ly$\alpha$ transition is saturated, complicating the measurement of absorbers with higher $N_{\rm HI}$ from Lyα only spectra. Similarly, for absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 15.5$, the Ly$\beta$ transition saturates. Thus, the precision and accuracy with which high-$N_{\rm HI}$ systems can be measured will depend to some degree on the number of accessible Lyman lines, thus imposing some redshift dependence.

For the sample presented in this paper, the QSO spectra cover the Lyα and Lyβ transition for all the absorbers. Given the typical spectral coverage of our sample (see Table 1), for absorbers with $z \gtrsim 2.3$, the Lyγ transition is also covered, and for those with $z \gtrsim 2.4$, Ly$\alpha,\beta,\gamma,\delta$ are generally observed. For absorbers with $z \gtrsim 2.5$, the spectral coverage of the QSOs typically allows for observation of transitions all the way to the Lyman Limit, thus allowing for accurate measures of $N_{\rm HI}$ to $\log(N_{\rm HI}/{\rm cm}^{-2}) \lesssim 17.2$.

An additional complication arises from the fact that high-$N_{\rm HI}$ absorbers are often found in blended systems of several high-$N_{\rm HI}$ absorbers. Such absorbers, when observed in only Lyα,β, often cannot be de-blended. The incidence of such blends increases with increasing redshift as the density of the forest increases. This blending effect is mitigated in our sample at $z > 2.4$ due to access to addition Lyman series lines, but may go unrecognized in the low end of the redshift interval ($z \lesssim 2.4$) which lack optically thin Lyman lines. For such blended systems in the presented sample with $z \lesssim 2.4$, the number of subcomponents may be underestimated which may lead to some degree of incompleteness in the sample.

In this appendix we assess the accuracy of our fits to absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 15.5$.

A.1. *Examples of fits to high-$N_{\rm HI}$ absorbers*

Figures 16 - 18 show examples of the fits to the HIRES spectra for absorbers with $15.5 < (\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.0$. In each case, the HIRES data are shown in black, and the solid red curve shows the model with the best fit to the data. The dashed curves show the $\pm 3\sigma$ formal errors on $N_{\rm HI}$ as reported by VPFIT. As expected, the errors on $N_{\rm HI}$ are dependent on the number of transitions observed; however, the dependence is not trivial as the effect of (a) blending from other proximate absorbers, (b) the contamination of higher-order Lyman transitions due to unrelated absorbers, and (c) the S/N of the spectrum in the areas of interest all affect our ability to accurately determine the parameters of the absorber.

A.2. *Testing fits with only Lyα and Lyβ*

To assess the accuracy of fits to absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 15.5$ observed only in Lyα and Lyβ, we use the set of absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) \gtrsim 15.5$ for which $n \geq 5$ Lyman transitions were measured, generally drawn from the subsample with $z > 2.4$. We then attempt to refit these absorbers using VPFIT with constraints from the Lyα and Lyβ region only. The outcomes of these tests can be generalized into two principal categories: highly-blended and less-blended absorbers.

Absorbers for which a portion of one of both edges/wings of the absorption line can be observed in Lyα and/or Lyβ can be fit much more reliably. Such systems, when refit with VPFIT using only their Lyα and Lyβ transition result in weaker constraints on the line parameters as expected, but result in $N_{\rm HI}$ determinations consistent within their formal errors with those obtained via the higher-order transitions. The fits are not systematically offset to higher or lower values of $N_{\rm HI}$, and therefor should not *bias* the measurements of $f(N,X)$ in any large way.

For absorbers whose Lyα and Lyβ transition are fully saturated and both edges/wings of the absorption line cannot be observed due to other neighboring absorbers (e.g. see the left-most panel in Figure 17 and the center panel in Figure 18), essentially no information is available on the sub-component structure of the system. Such systems are often fit with a minimum of two components because it is rare that the wings/edges of the blended system are symmetric. In the absence of higher-order Lyman lines, the number of subcomponents is typically underestimated, likely leading to some degree of incompleteness in the lower-redshift portion of the catalog. There would also be a tendency to underestimate the total $N_{\rm HI}$ in such blends. If the $N_{\rm HI}$ of individual absorbers in the blends were systematically underestimated, this would *bias* the measurement of $f(N,X)$ to have steeper values of β than the true value.

In conclusion, the lack of higher-order constraints on the lower-z saturated absorbers have two principal effects on our measurements. (1) Relatively unblended absorbers yield somewhat weaker constraints on the parameters that are unlikely to bias the measurement of $f(N,X)$. (2) Highly-blended absorbers yield very weak constraints from Lyα and Lyβ only and therefore are typically fit with fewer subcomponents and with lower-$N_{\rm HI}$ values than would be assigned using additional Lyman series constraints. This results in some degree of incompleteness in the lower-z portion of the catalog.

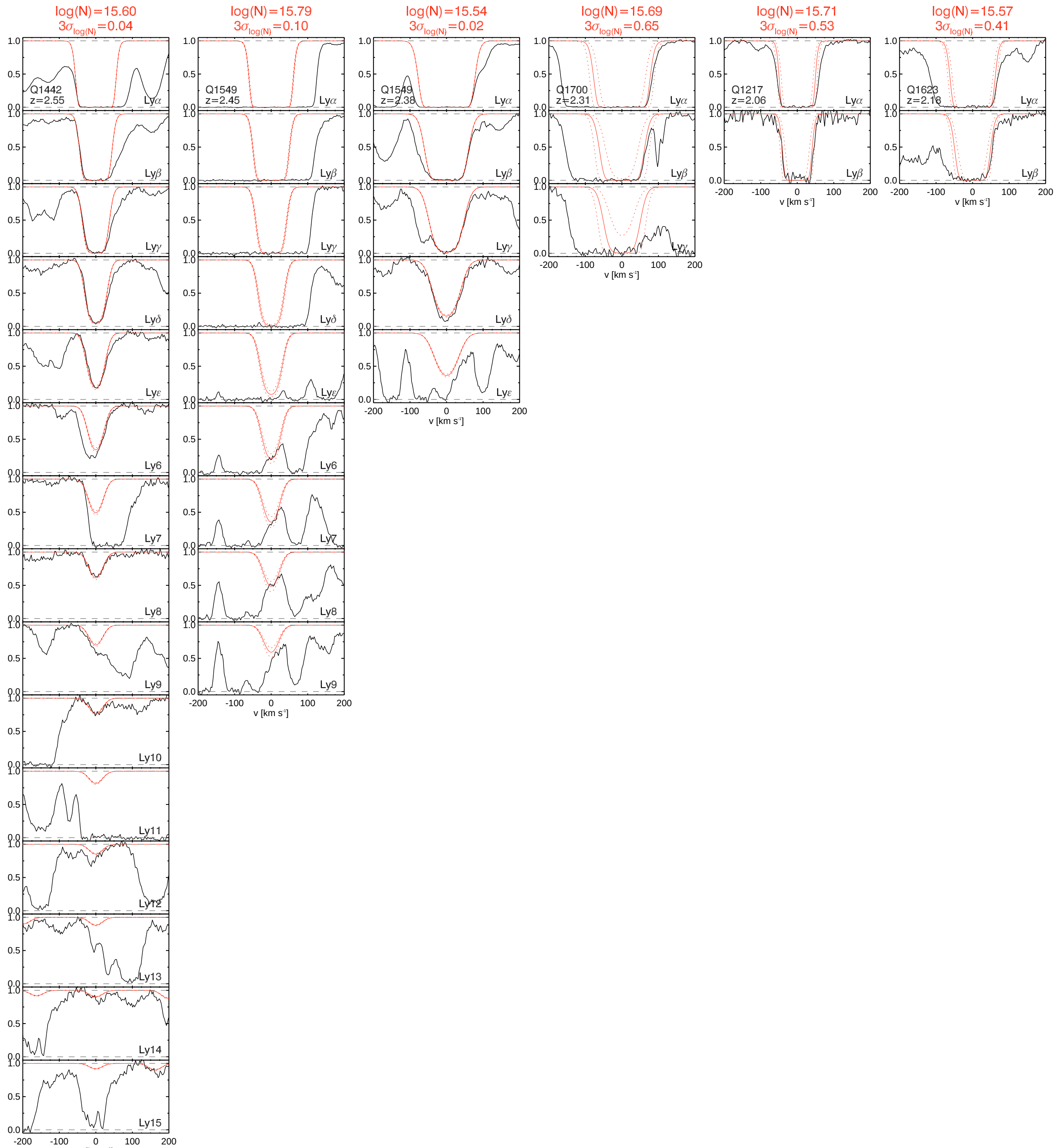

Figure 16. Example fits to absorbers with $15.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 16.0$. The black curves show the HIRES data. Grey dashed lines mark the zero point and continuum of the QSO spectra. The solid red curve shows the best fit Voigt profile to each absorber. The dashed red curves show the formal $\pm 3\sigma$ $N_{\rm HI}$ error.

A.3. *Assessing Incompleteness in the $z \lesssim 2.4$, log($N_{\rm HI}/{\rm cm}^{-2}$) > 15.5 Catalog*

To estimate the degree of potential incompleteness in the lower-z portion of the catalog due to line blending, we consider the fraction of the higher-z, high-$N_{\rm HI}$ absorbers observed in $n \geq 5$ Lyman series transitions that would be unconstrained or miscounted with observations of Lyα and Lyβ alone. This method likely overestimates the incompleteness as the degree of line blending and contamination is higher in the high-redshift forest than in the lower-z forest due to evolution in the line density as a function of redshift. Nevertheless, we report the results of this test, considering separately absorbers in half-dex bins of $N_{\rm HI}$. The results are summarized in Table 9.

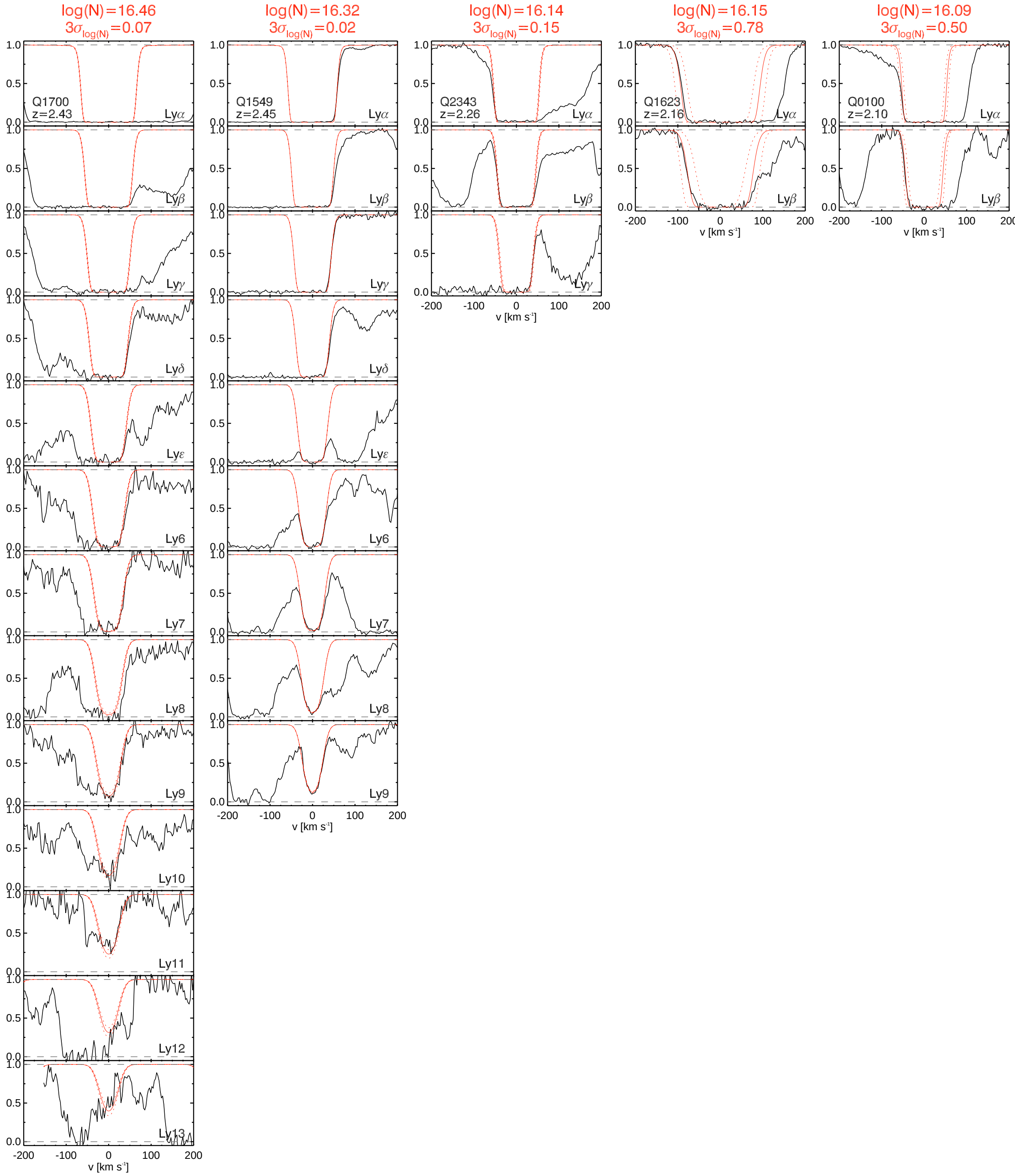

Figure 17. Example fits to absorbers with $16.0 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 16.5$. The black curves show the HIRES data. Grey dashed lines mark the zero point and continuum of the QSO spectra. The solid red curve shows the best fit Voigt profile to each absorber. The dashed red curves show the formal $\pm 3\sigma$ $N_{\rm HI}$ error.

For absorbers with $15.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 16.0$, 23 are measured with 5 or more Lyman series transitions. Of these, only 2 would have essentially no constraint on the parameters of the absorber from Lyα and Lyβ alone (e.g. Figure 16, second from the left). For 7/23 of these systems, the constraint on $N_{\rm HI}$ would be relatively weak, however, an individual component associated with that absorber can be clearly identified. All 7 of these systems are fit with much improved accuracy with the additional constraint from Lyγ. Among absorbers with $16.0 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 16.5$, we observed 12 absorbers with constraints from 5 or more Lyman series transitions. Of these, only 4 are strongly constrained from Lyα and Lyβ alone. With Lyγ, 8/12 could be fit, and with Ly$\alpha,\beta,\gamma,\delta$, 11/12 would be strongly constrained and could accurately be fit. For absorbers with $N_{\rm HI}$ just below that of LLSs ($16.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.0$), we observe 5 systems in 5 or more Lyman series transitions.[16] Of these, only 2 would

[16] All five absorbers are observed in a minimum of 9 transitions. We do not observe any absorbers in this $N_{\rm HI}$ range with 5-8 Lyman series transitions observed.

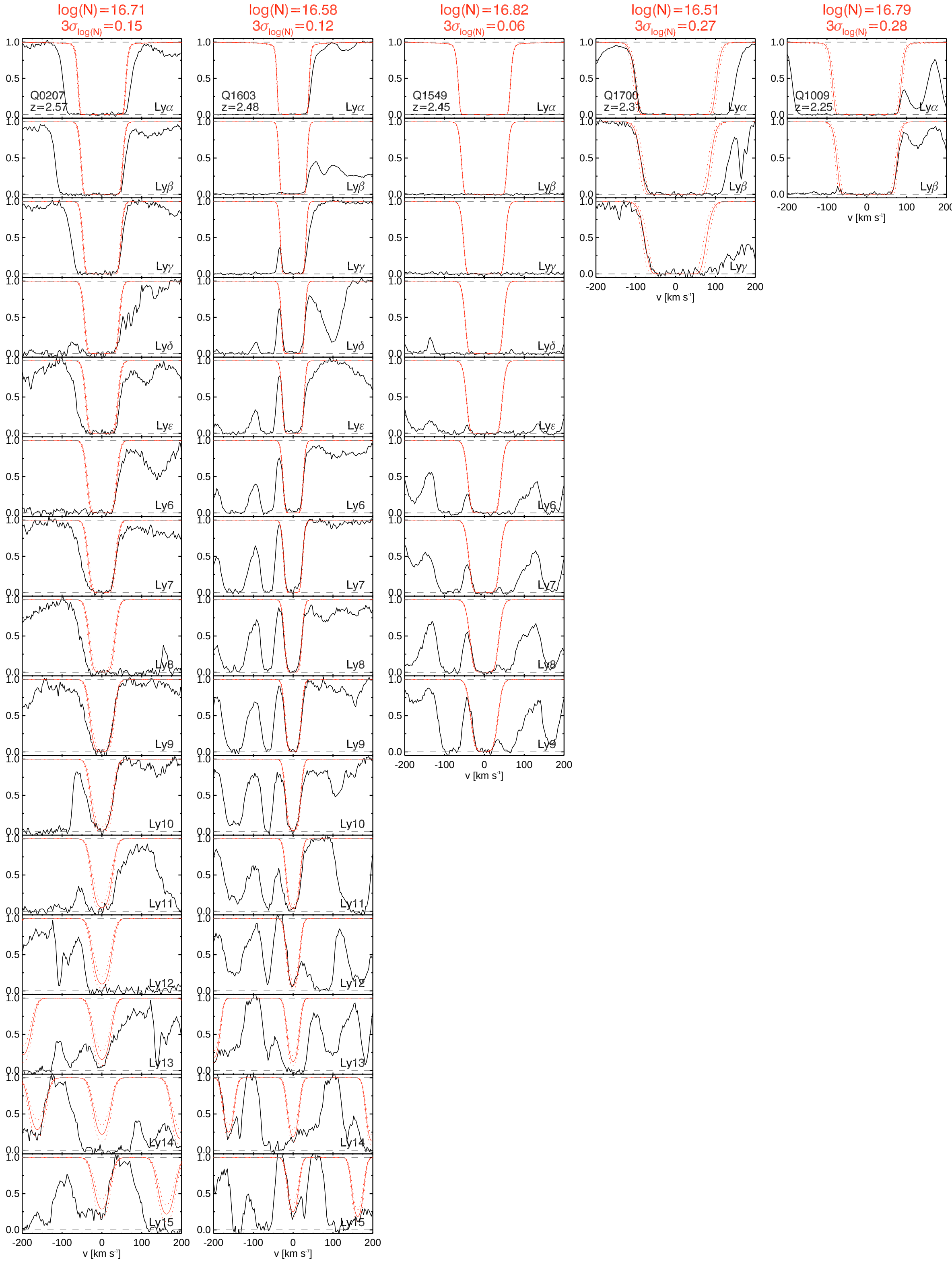

Figure 18. Example fits to absorbers with $16.5 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 17.0$. The black curves show the HIRES data. Grey dashed lines mark the zero point and continuum of the QSO spectra. The solid red curve shows the best fit Voigt profile to each absorber. The dashed red curves show the formal $\pm 3\sigma$ $N_{\rm HI}$ error.

Table 9
Constraints on $N_{\rm HI}$ vs. the number of observed Lyman series transitions

$\log(N_{\rm HI}/{\rm cm}^{-2})$	Lines considered in the fit	# with 5+ Lyman series lines[a]	strongly constrained #	strongly constrained %	weakly constrained[b] #	weakly constrained[b] %	unconstrained[c] #	unconstrained[c] %
15.5 - 16.0	Lyα,β	23	14	61%	7	30%	2	9%
15.5 - 16.0	Lyα,β,γ	23	18	78%	3	13%	2	9%
16.0 - 16.5	Lyα,β	12	2	17%	2	17%	8	67%
16.0 - 16.5	Lyα,β,γ	12	2	17%	6	50%	4	33%
16.5 - 17.0	Lyα,β	5	1	20%	2	40%	2	40%
16.5 - 17.0	Lyα,β,γ	5	3	60%	0	0%	2	40%

[a] The number of absorbers in the catalog with $N_{\rm HI}$ in the range listed and at least 5 Lyman series transitions observed.
[b] For weakly constrained absorbers, the transitions listed give unambiguous evidence of the presence of an absorber, and typically have at minimum one edge of one transition well observed without a significant blend.
[c] For unconstrained absorbers, the transitions listed are fully saturated and blended with other high-$N_{\rm HI}$ lines (e.g. see middle panel Figure 18) meaning this absorber would likely not be counted or be very poorly fit in data with only the transitions listed.

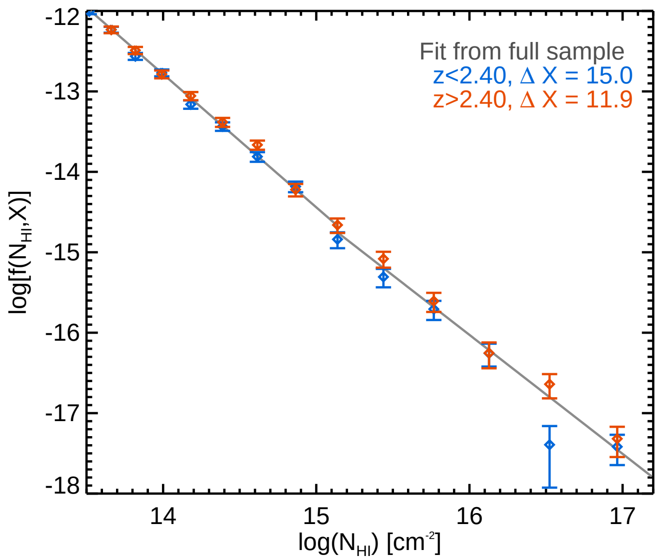

Figure 19. The frequency distribution measured from the higher-redshift (red) and lower-redshift (blue) portions of the absorber sample (many of the points from the two subsamples lie on top of each other). The grey curve shows the broken power law fit to the full absorbers sample from §5.1.

have no constraint on $N_{\rm HI}$ from Lyα and Lyβ alone.

From this assessment based on higher-z measurements, we expect that the principal bias in our measurements of $N_{\rm HI}$ for high-$N_{\rm HI}$ systems for absorbers with only Lyα and Lyβ is to under-count the number of systems with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 15.5$. Within the present sample, systems with $16.0 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 16.5$ appear to be most-affected by blending and may therefore be the most incomplete. In the next appendix, we add further evidence to this claim by comparing the measurement of $f(N,X)$ for absorbers in two redshift subsamples.

B. THE EFFECT OF THE ACCURACY OF $F(N,X)$ ON THE MEASUREMENT OF THE $\lambda_{\rm mfp}$

Here we assess the redshift dependence of the frequency distribution presented in §3 and §5. We split the absorber sample roughly in half and compare those absorbers with $z < 2.4$ and those with $z > 2.4$. For the higher-redshift subsample, the spectra typically cover at least 4 Lyman series transitions and most of this sample have all the Lyman series transitions observed. Figure 19 shows the results of this comparison. Notably, the higher-redshift subsample (red points) is in reasonable agreement with the broken power law fit (gray curves) presented in §5.1 that were used in the Monte Carlo simulation in §6. Further, the higher-redshift sample, tends to lie slightly above the fit. Some portion of this is likely due to the redshift evolution of the line density of the forest; however, it is also possible that $f(N,X)$ of absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 15.5$ is actually higher than that measured from the full sample. As discussed in Appendix A.3, we expect that the lower-z catalog likely suffers from incompleteness due to line blending, especially for absorbers with $16.0 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 16.5$. In Figure 19, we note the only bin for which the low-z and high-z data do not agree within 1σ contains absorbers with $16.0 < \log(N_{\rm HI}/{\rm cm}^{-2}) < 16.5$.

The principal result of this paper is the relatively high measured value of $f(N,X)$ for absorbers with $14.5 < (\log(N_{\rm HI}/{\rm cm}^{-2}) < 17.2$ which results in a value of $\lambda_{\rm mfp}$ comparable to or smaller than previous measurements. If we were to use the high-redshift sub-sample from Figure 19 to measure the opacity and $\lambda_{\rm mfp}$, we would find higher IGM opacity and therefore a shorter (more discrepant) $\lambda_{\rm mfp}$. In any case, if there is a systematic bias in the measurements, it is one that would conspire to make $\lambda_{\rm mfp}$ appear

larger than its true value. We therefore conclude that it is unlikely that a bias in the precision of our measurements of $f(N,X)$ as a function of z would change the principal result of the paper.